DESIGN OF A FLEXIBLE AND LOW-POWER IONOSPHERIC SOUNDER

By

Alex Morris

RECOMMENDED:

Dr. Joseph Hawkins

Dr. Vikas Sonwalkar

Dr. William Bristow
Advisory Committee Chair

Dr. Charles Mayer
Chair, Department of Electrical and Computer Engineering

APPROVED:

Dr. Douglas Goering
Dean, College of Engineering and Mines

Dr. John Eichelberger
Dean of the Graduate School

4/24/14

Date

DESIGN OF A FLEXIBLE AND LOW-POWER IONOSPHERIC SOUNDER

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Alex Morris, B.S.

Fairbanks, Alaska

May 2014

Abstract

Characterizing the structure of the ionosphere has practical applications for telecommunications and scientific applications for studies of the near-Earth space environment. Among several methods for measuring parameters of the ionosphere is ionospheric sounding, a radar technique that determines the electron content of the ionosphere as a function of height. Various research, military, and commercial institutions operate hundreds of ground-based ionosondes throughout the globe, and new ionosondes continue to be deployed in increasingly remote and distant locations. This thesis presents the design of an ionospheric sounder that reduces the power, size, and cost compared to existing systems. Key improvements include the use of an open-source software-defined radio platform and channel-aware dynamic sounding scheduling.

Table of Contents

Page

List of Figures

List of Tables

Page

Chapter 1 Introduction

1.1 A Brief Historical Background

Perhaps the oldest observations of the Earth's upper atmosphere are of the aurora, a beautiful phenomenon that has existed for longer than mankind. Only in the 18th century did astronomers start to notice the correlation between the frequency of aurora and the 11-year period of the sunspot cycle, but with vision being the only way to observe the near-Earth space environment, scientists had a limited understanding of the system's mechanics.

The turn of the 20th century saw rapid improvements in electronics and long-distance communications, providing both the tools and motivation to study the upper atmosphere. Communications in the late 1800's and early 1900's progressed from short telegraph links to long telegraph networks and soon to wireless telegraphy by radio. Gugliemo Marconi achieved the first undisputed transatlantic radio communications in 1902, a feat that would later be explained by properties of the ionosphere. Marconi quickly began to compete against the established wired telegraphy companies as wireless radio required much less infrastructure than its wired counterpart. However, there was much greater variability in its signal quality. Signals were much weaker in the daytime,

only signals below a certain frequency could be heard, and there were other inexplainable phenomena. The commercial importance of wireless telegraphy provided the motivation for studying the Earth's upper atmosphere, and researchers had the radio as a new tool.

The Brit Oliver Heaviside and the American Arthur Kennelly independently and almost simultaneously in 1902 proposed the existence of an ionized layer of gas in Earth's atmosphere, now known as the Kennelly-Heaviside layer, to explain the ability of Marconi's transatlantic signals to travel so far about Earth's surface. The idea of skywave propagation, where a radio signal is reflected between the ionosphere and the Earth over long distances, was the generally accepted theory in the early 1900's, but it wasn't until 1924 that Edward Appleton provided hard evidence of the ionosphere's existence by performing many radio experiments, including operating a pulsed radar system which was the first ever ionosonde. Appleton devoted his life to experiments and studies of the ionosphere, and developed a magneto-ionic theory that set up much of the framework for mathematical analysis of waves in the ionosphere.

The First and Second World Wars drove much of the research in ionospheric physics in the early and mid 1900's, where predicting channel conditions for skywave communications and designing appropriate communications infrastructure were paramount. After the World War II, the importance of skywave communication gradually diminished as wired communication infrastructure steadily improved in

developed nations, but skywave communication remained and will likely forever remain important for communications in marine, aeronautical, remote site, emergency, amateur, and broadcast applications.

The invention of Earth-orbiting satellites and other space borne systems in the mid 1900's marked a new era in ionospheric research. Satellites exist in an imperfect environment. Communications between space and ground must penetrate the ionosphere, modifying the propagating electromagnetic wave by diffraction, Faraday rotation, and absorption. Bulges in the ionosphere caused by tidal or heating effects increase drag on satellites in low-earth orbit. Radiation from the Van Allen belts, which is not strictly an ionospheric process, can cause damage to satellites. All of the above effects can be understood by studying the ionosphere, so in the interest of designing better space systems there is motivation for ionospheric research. Space systems have also provided new tools for observing the regions near Earth. Soundings of the ionosphere above the F-region peak, measurements of the magnetic field in Earth's magnetosphere, and direct observation of solar wind have all been made possible by spacecraft. As a result, we have a much broader knowledge of the upper atmosphere and regions of space near Earth.

Today there is a wealth of experimental and observational information for researchers, and new research continues. The technological implications of the

ionosphere remain important, and many interesting scientific questions about the ionosphere remain unanswered.

1.2 The Ionosphere

The ionosphere is a region in the Earth's atmosphere at altitudes from approximately 50 km to 1000 km where the concentration of ions is significant compared to the neutral atmosphere, with a peak ion concentration usually occurring somewhere between 200 km and 500 km altitude. The primary source of ionization is solar ultraviolet radiation and to a much lesser extent cosmic radiation. The different radiation spectra and the various species and molecular densities in the atmosphere result in layers of the ionosphere that are characterized by their effects on radio signals and distinguished by their local peak electron densities:

The D-region, from approximately 50 km to 100 km altitude, occurs during daylight hours and then rapidly dissipates at night. The atmospheric density is relatively high at these lower altitudes of the ionosphere, and so ion-neutral collisions are relatively frequent. This is the reason for attenuation of shortwaves signals during daytime hours.

The E-region exists from approximately 100 km to 200 km altitude and largely dissipates at night. Like the D-region, the E-region can also cause significant attenuation

of MF and HF radio waves due to frequent ion-neutral collisions. A second type of E-region layer called the sporadic E (Es), and whose causes are less well understood, exhibits extraordinarily high plasma density in thin regions for times lasting only for a few minutes to several hours. The existence of sporadic E can enhance long distance radio communications of frequencies as high as the UHF range.

The F-region, from 200 km altitude to 1000 km, is the highest altitude and usually the most dense of the ionospheric layers. It exists during day and night and is the main element in providing skywave communication for HF links. The F-region is often characterized the maximum frequency of an incident electromagnetic wave that will be reflected, the peak ion density, and the altitude of that peak density.

To a very rough approximation, the ionosphere may be viewed as a spherical shell surrounding the Earth, which explains the ability of shortwave radio signals to propagate such long distances. However, to more fully characterize the ionosphere and explain more interesting phenomena, the spatial and temporal variations in the ionosphere must be explored.

The motion of the Earth relative to the sun causes variation in the intensity of solar radiation and explains much of the diurnal, seasonal, and geographic variation in the structure of the ionosphere. During the night the rate of recombination of ions dominates the rate of ionization of neutral particles and the ionosphere tends to drop in density and

move to higher altitudes compared to the daytime ionosphere. A similar effect occurs due to seasonal and geographic variation: the lower solar radiation of the wintertime and high latitudes tend to result in the ionosphere being lower in density and higher in peak altitude compared to the ionosphere during summertime and at lower latitudes.

There are also anomalous but predictable variations in the structure of the ionosphere. The winter anomaly is the effect of relatively lower ion concentration in the summer time when the summer ionization is expected to be greater than winter ionization. The anomaly occurs primarily in the Northern hemisphere and is due to a seasonal change in chemical composition in the upper atmosphere. The equatorial anomaly is a an increased ion concentration at latitudes +/- 20 degrees from the magnetic equator and is caused by the combined effects of solar heating, tidal motion in the atmosphere, and latitudinal mobility of ions along the magnetic field lines.

Finally, sporadic solar events will cause variations to the structure of the ionosphere. Polar cap absorption (PCA) events are the occurrence of unusually large fluxes of high-energy protons emitted by the sun, enhancing the D-region ionization at high latitudes and making HF propagation across the polar caps extremely difficult. Similarly, sudden ionospheric disturbances (SID's) associated with increased solar x-ray radiation enhance the dayside D-region. Geomagnetic storms cause increased precipitation of particles into the upper atmosphere in the auroral ovals, enhancing the

ionospheric density in these regions.

The ionosphere is largely influenced by solar effects, but interaction with the Earth's geomagnetic field and magnetosphere are also important. Since the ionosphere is composed of charged particles, it affects and is affected by the ambient electric or magnetic fields surrounding Earth. The ionospheric dynamo current, caused by plasma motion across magnetic field lines and an associated charge separation and electric field, depletes the Earth's magnetic field strength on the day-side. The tangential drag on the outer shells of the magnetosphere carry magnetic field lines anti-sunward across the polar cap, and field lines return to the day-side at lower latitudes. The moving field lines have an associated electric field to an observer on earth and a large cross-polar-cap potential from the dawn to the dusk side of the Earth. The resulting field-aligned current system is complicated by the anisotropic properties of the magnetized plasma and spatial variation of the relative ion and neutral densities in the ionospheric plasma. Currents exist both parallel and perpendicular to the dawn-to-dusk electric field, and sometimes very large currents can create large deviations in the geomagnetic field on Earth's surface and even induce electric currents in pipelines and electric transmission lines.

The exact structure and behavior of the ionosphere depend on the many factors described above. General predictions are possible based on climatological models such as the International Reference Ionosphere (IRI), although they are unreliable compared to

direct local observations. Some of the challenges in studies of the ionosphere include improving the predictive power of mathematical models and understanding the mechanisms of ionospheric behavior.

1.3 Instruments for Studying the Ionosphere

Of the many tools used to study the ionosphere, the majority involve some kind of remote sensing technique -- measuring the reflection or refraction of incident electromagnetic waves from or through the ionospheric plasma. The ionosphere's effect on electromagnetic waves may yield information such as the plasma density, location, and/or motion.

Ionosondes detect reflections that are essentially unaltered versions of the transmitted pulse since reflections are specular and the target is large and uniform compared to the operating wavelength of the incident wave. The ionosphere behaves as a mirror to HF waves below the F-region critical frequency and essentially all incident power is reflected. Ionosondes are relatively inexpensive and provide information about the electron content of the ionosphere as a function of height for regions below the F region peak. Ionosondes may also be operated on board satellites, where they may view the topside of the ionosphere above the F region peak and are thus known as topside

sounders.

The SuperDARN network of radars exploit Bragg-scatter power reflected from field-aligned irregularities (FAI's) in the ionosphere to measure horizontal plasma velocities across an enormous field of view. In the high latitude regions of the Earth, magnetic field lines extend in to the outer magnetosphere where there is interaction with the solar wind. The plasma's high mobility parallel to magnetic fields results in field lines being equipotentials. Hence, plasma motions in the outer magnetosphere are communicated along field lines into the ionosphere, where SuperDARN observes them.

Incoherent scatter radars (ISR's) measure the backscatter power from small density fluctuations in the ionosphere. ISR's are typically composed of large, high power antenna arrays or a single high-gain antenna and operate at frequencies of hundreds of MHz. ISR's provide a wealth of information about the local ionosphere, including composition, ion and electron temperatures, and density. They are comparatively expensive to build and operate.

Several techniques do not employ remote sensing, but instead rely on in-situ measurement. These include rockets that probe the ionosphere by emitting radio signals and seeking resonances with the immediately-surrounding plasma, and/or using particle detectors to sample the composition of the atmosphere directly. Also included in this category are low-earth orbiting satellites that are equipped with chemical samplers,

magnetometers, and electric field probes.

Long-term trends in technology and infrastructure have fueled interests in studies of the ionosphere. A growing technological application of ionospheric sciences is for Earth-orbiting satellites. Geomagnetic storms and other anomalies in the structure of the ionosphere can disrupt the trajectory of satellites and sometimes even cause damage, so predicting and characterizing these anomalies is beneficial to the operations of these satellites. Despite the relatively high radio frequencies used by satellites to communicate with ground stations, the ionospheric plasma can significantly alter the propagating signal so characterizing the structure of the ionosphere is again beneficial to satellite operations. And finally, even though HF communications have become less important in everyday life as other technologies have superseded it, HF will always be an important radio band for remote and mobile radio links, military communications, and emergency services. Studies of the ionosphere will remain important for these applications.

Higher quality observations and greater breadth of observations improve the understanding of the ionosphere's structure and behavior. This can be done by designing smaller, less expensive, and more efficient instruments. This thesis presents ideas to help achieve that end.

1.4 Thesis Organization

This thesis is organized to parallel the phases of design of an ionosonde. Background information is given in Chapter 2 and Chapter 3, with Chapter 2 discussing the general theory of electromagnetic waves in the ionosphere and Chapter 3 discussing general radar techniques. Chapter 4 follows with a survey of two modern ionospheric sounders. Chapter 5 describes the design of an ionospheric sounder that is based on the topics in the previous three chapters – general theory and case studies – yet tries to improve upon previous technology. Chapter 6 describes the results and evaluates the performance of the new system, and Chapter 7 concludes with a discussion of possible applications and ideas for future improvements to the ionosonde design.

Chapter 2 Radio Waves and the Ionosphere

2.1 Dispersion Relation of Electromagnetic Waves in the Ionosphere

The analysis of electromagnetic waves in the ionosphere begins with an analysis of the plasma medium. The assumption of a ideal plasma is valid by measure of the plasma parameter:

$$\Lambda = 4\pi n_e^{-1/2} \left(\frac{\epsilon_0 k T_e}{q_e^2} \right)^{3/2} , \tag{2.1}$$

where n_e is the electron number density, ε_0 is the permittivity of free space, k is Boltzmann's constant, T_e is the electron temperature, and q_e is the charge of a single electron. For values of Λ much greater than unity, the effects of collisions and electric force between particles are small compared to their thermal energy. At altitudes of 200 km a typical value of the plasma parameter is 10^5, showing that the ideal plasma assumption is valid. At 50 km altitude a typical value is 10^2, showing that the ideal plasma assumption is still valid, although the non-ideal effects of the plasma are more significant [Chen, 2006].

A wave incident on a plasma will excite motion in the electrons (and a small, often negligible motion in the ions) by electromotive force, and this electron motion will

modify the wave behavior. When the plasma is magnetized and there is a component of the electric field that is perpendicular to the magnetic field direction, as in most cases of the ionospheric plasma, then the motion of the electrons is complicated due to the Lorentz force ***F***:

$$\vec{F} = q\left(\vec{E} + \vec{v}\, x\, \vec{B}\right) \quad , \tag{2.2}$$

where q is the charge of an electron, ***E*** is the electric field, ***v*** is the velocity, and ***B*** is the ambient magnetic field. The magnetic field of the incident wave also modifies the electron motion via the Lorentz force, but its effects are assumed to be small compared to the effects of the ambient magnetic field. The ambient magnetic field causes an acceleration of the electrons that has components both parallel to and perpendicular to the incident wave electric field. Thus the magnetized plasma may be described as anisotropic because the wave behavior in a magnetized plasma depends on the direction of the propagating wave and also gyrotropic because secondary field components may exist that are orthogonal to the incident field. The material properties of the plasma medium such as impedance and permittivity have cylindrical symmetry about the magnetic field axis, and they are expressed as a tensor with at least five elements rather than a simple scalar or a diagonal matrix.

Expressions for waves in a plasma are found by linearizing the equations of motion (Newton's equations, the Lorentz force, and Maxwell's equations) and assuming a

small perturbation in the electric field. The system of equations reduce to a set of solutions, the choice of which solution depends on the polarization and propagation direction of the wave relative to the magnetized plasma. The wave behavior is a function of the material properties of the plasma and the frequency of the incident wave. It is often expressed as a dispersion relation in terms of n^2:

$$n^2 = \frac{c^2 k^2}{\omega^2} \quad , \tag{2.3}$$

where c is the speed of light in free space, k is the wavenumber, and ω is the wave frequency. The value of n is known as the refractive index of the medium and is the ratio of phase velocity of the wave in free space to the phase velocity of the wave in the plasma medium. The expression for n^2 of the material describes its refractive properties, dispersive effects, phase and group velocities, and conditions for reflection or resonance.

Using the assumption that the ion motion is negligible compared to the electron motion, the Appleton-Hartree equation is a general solution to the dispersion relation of a wave in a magnetized plasma. It expresses the refractive index in terms of four items: the plasma frequency ω_N, the electron gyrofrequencies ω_{eL} and ω_{eT} about the components of the magnetic field that are longitudinal ($\mathbf{B_T}$) and transverse ($\mathbf{B_T}$) to the propagation vector, and the effective collision frequency of the plasma ν:

$$n^2 = \frac{X}{1 - iZ - \frac{Y_T^2}{2(1 - X - iZ)} \pm \left(\frac{Y_T^4}{4(1 - X - iZ)^2} + Y_L^2 \right)^{1/2}} \quad , \tag{2.5}$$

where

$$X = \frac{\omega_N^2}{\omega^2} \quad , \tag{2.6}$$

which is the square of the ratio of plasma frequency to wave frequency,

$$Y_L = \frac{\omega_{eL}}{\omega} \quad , \tag{2.7}$$

which is the ratio of gyrofrequency about the longitudinal component of the magnetic field to the wave frequency,

$$Y_T = \frac{\omega_{eT}}{\omega} \quad , \tag{2.8}$$

which is the ratio of gyrofrequency about the transverse component of the magnetic field to the wave frequency, and

$$Z = \frac{\nu}{\omega} \quad , \tag{2.9}$$

which is the ratio of the effective collision frequency to the wave frequency. The +/- sign in the denominator indicates two modes of propagation in a magnetized plasma that are distinguished by their different dispersion relations and different polarizations.

A wave (or superposition of the two wave modes) launched vertically into the ionosphere will encounter a plasma of increasing density, and the index of refraction for that wave mode will change according to the local plasma density. The propagation of

the wave changes according to the changes in index of refraction and in special cases the wave will no longer propagate upward. If the index of refraction for the upward-traveling wave becomes infinite then the wave will be absorbed in the resonant plasma. If the index of refraction transitions from positive to negative then the upward-traveling wave will become evanescent and all wave power is reflected downward.

Ignoring the effects of collisions and considering a few special geometries for the Appleton-Hartree equation illustrates some basic behavior of waves in a plasma. For completely longitudinal propagation (***k*** parallel to ***B***), the two wave modes correspond to left-hand and right-hand circularly polarized waves. Their dispersion relations are given by:

$$n^2 = 1 - \frac{X}{1 \pm Y} \quad . \tag{2.10}$$

The zeros mark the conditions for reflections, and any poles mark conditions for resonance. The dispersion relations indicates two reflection points at frequencies marked by

$$X = 1 \pm Y \quad , \tag{2.11}$$

and resonance at

$$Y = 1 \quad . \tag{2.12}$$

The following figure illustrates the dispersion relation for the two wave modes.

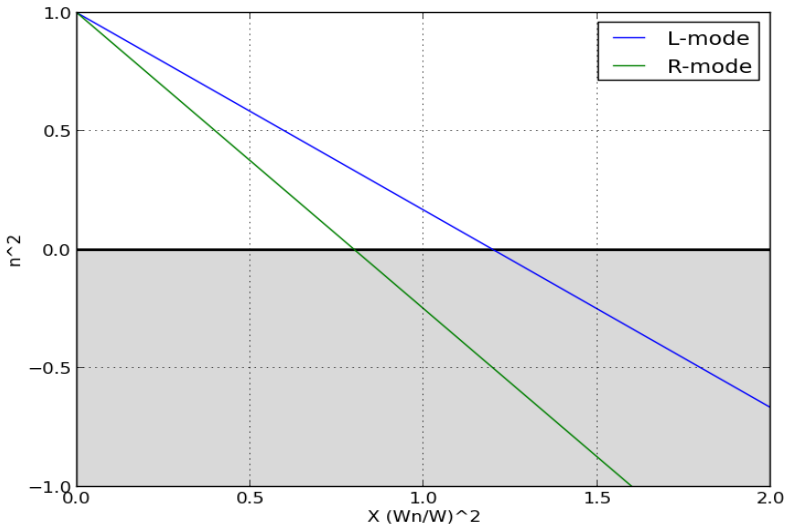

*Figure 1: Dispersion relation of EM wave in a plasma, **k** parallel to **B***

For complete transverse propagation (***k*** perpendicular to ***B***), the two modes correspond to a linearly-polarized "ordinary" wave (***E*** parallel to ***B***) whose dispersion relation is unaffected by the magnetizing field and a linearly-polarized "extraordinary" wave (**E** perpendicular to **B**) that is modified by the magnetizing field. For the ordinary wave,

$$n^2 = 1 - X \quad , \tag{2.13}$$

with reflection at

$$X = 1 \quad . \tag{2.14}$$

For the extraordinary wave,

$$n^2 = 1 - \frac{X(1-X)}{1-X-Y^2} \quad , \tag{2.15}$$

with reflections at

$$X = 1 \pm Y \quad , \tag{2.16}$$

and resonance at

$$X = 1 - Y^2 \quad . \tag{2.17}$$

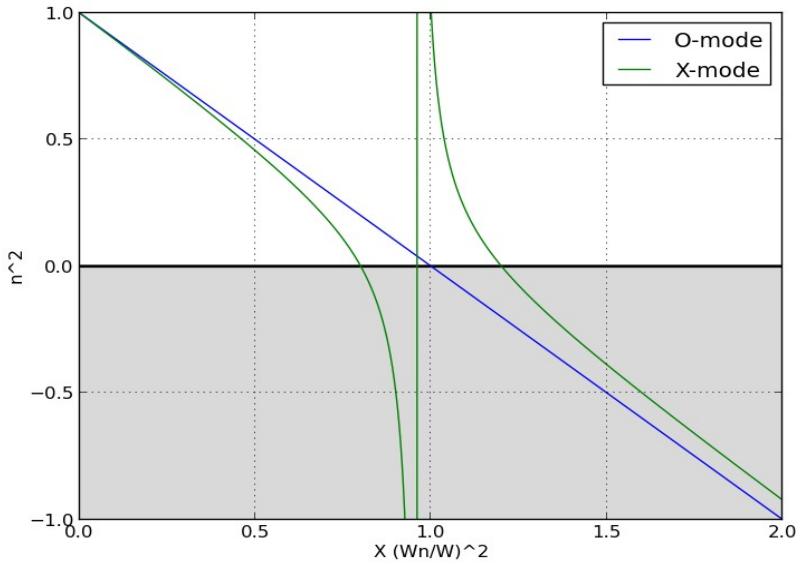

Figure 2: Dispersion relation of EM wave in a plasma, ***k*** *perpendicular to* ***B***

For almost all cases of ionospheric sounding there are components of the magnetic field that are both transverse and longitudinal to the wave propagation vector ***k***, and therefore the dispersion relation may be viewed as being somewhere between the

special cases of transverse and longitudinal propagation. The geometry of the apparatus is illustrated below.

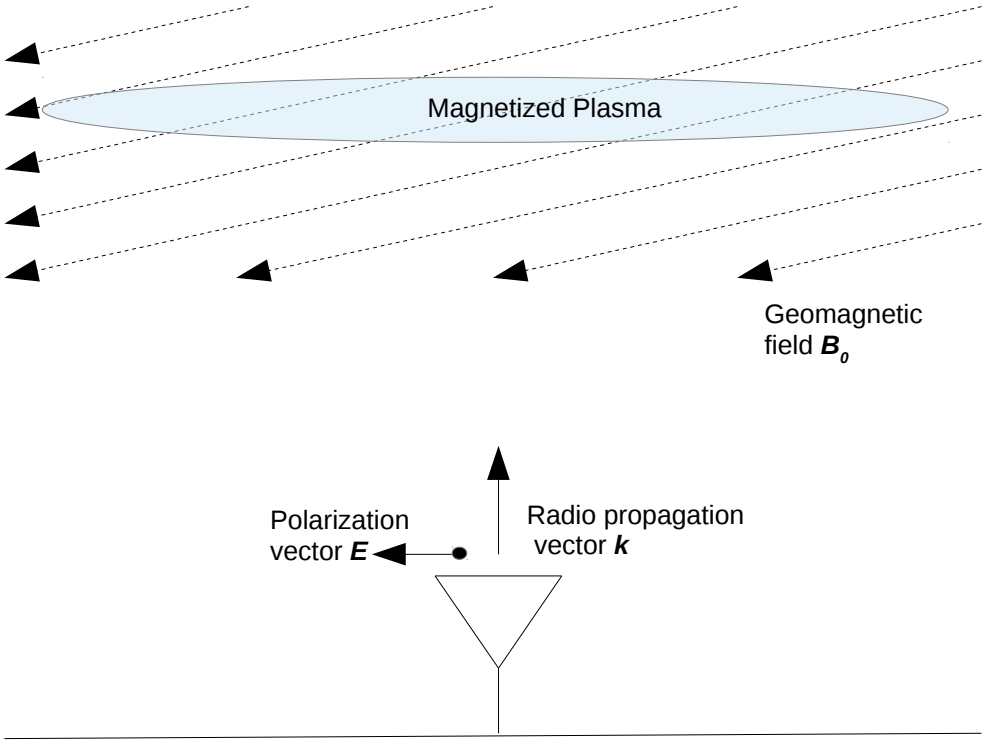

Figure 3: Diagram of the generic ionospheric sounder

Magnetic inclinations in mid-latitudes typically range from 60 degrees to 80 degrees, and so a vertically-incident wave has a significant longitudinal component to the magnetic field. The dispersion relation is illustrated below for 70 degrees magnetic inclination, a gyrofrequency of 1.5 MHz, and plasma frequency of 6 MHz.

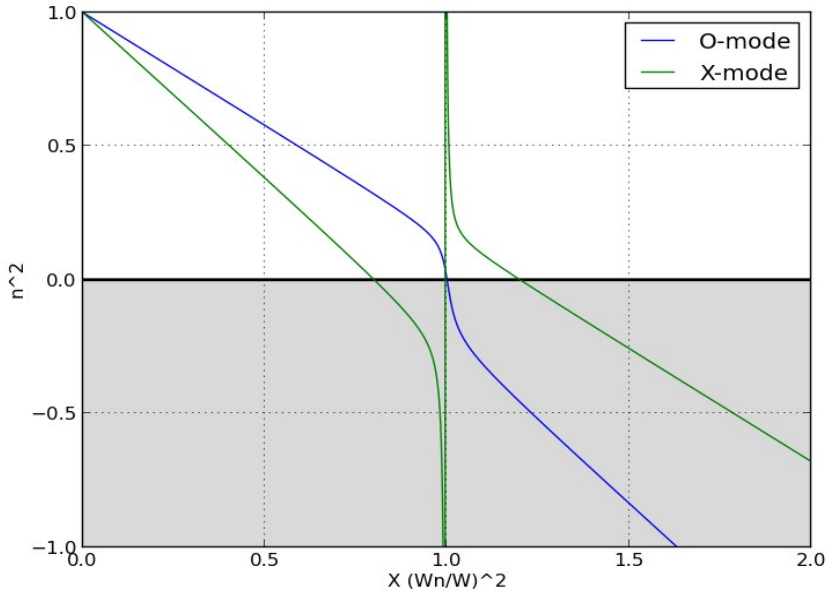

*Figure 4: Dispersion relation of EM wave in a plasma, **k** oblique to **B***

A wave with arbitrary polarization launched from the ground may be viewed as some superposition of the two elliptically-polarized wave modes. When a linearly polarized wave, which consists of two counter-rotating elliptical polarizations, is launched and enters the plasma medium, it encounters an increasing plasma density and the two characteristic polarizations diverge, following different propagation paths due to their differing indices of refraction. The reflection condition for the two waves is also different. For the ordinary wave the condition is the same as that for an unmagnetized plasma:

$$n^2 = 1 - X = 0 \quad , \tag{2.18}$$

or

$$\omega_N = \omega \quad . \tag{2.19}$$

For the extraordinary mode, the reflection condition is modified by an additional term involving the electron gyrofrequency:

$$n^2 = 1 - X - \sqrt{Y_L^2 + Y_T^2} = 0 \quad , \tag{2.20}$$

or

$$\omega_N \simeq \omega - \frac{\omega_e}{2} \quad . \tag{2.21}$$

Consequently, there appear two traces in a normal ionogram, with the x-mode being reflected at a lower height and its critical frequency being larger by approximately one half the gyrofrequency compared to the ordinary wave.

A third trace called the “z” mode sometimes exists, corresponding to the lower reflection condition of the extraordinary mode:

$$n^2 = 1 - X + \sqrt{Y_L^2 + Y_T^2} = 0 \quad , \tag{2.22}$$

or

$$\omega_N \simeq \omega + \frac{\omega_e}{2} \quad . \tag{2.23}$$

The z-mode rarely exists because in most soundings all of the incident wave power is reflected before this condition is met. However, at high latitudes the large magnetic inclination may allow for coupling of energy into waves that will reach these higher refection conditions. An exact description of the mechanism for this coupling is beyond

the scope of this text, but it involves a consideration of the collision frequency of the plasma and nearly-longitudinal wave propagation along the magnetic field direction [Rawer, 1993]. In topside sounding, where the sounder exists in a tenuous plasma, it is possible to directly excite the z-mode of propagation and these z-mode echoes are usually detected.

At frequencies below the electron gyrofrequency the right-hand circularly polarized wave mode exists and is longitudinal with the magnetic field direction. This is known as a whistler wave. Whistler wave-related reflections are very rarely seen in ground-based soundings since soundings are rarely done on these low frequencies and reflected energy must come via coupling with some other wave mode or irregularities in the plasma medium. Whistler wave-related reflections are more commonly studied in topside sounding

The index of refraction for a plasma is not equal to that of free space, and so the propagation of the wave is altered by the plasma even when the reflection condition is not (yet) met. The group velocity of the wave is slower in the plasma, which means that the "virtual" height of the ionosphere deduced by sounding will be greater than the true height. Also, oblique propagation through the magnetized plasma creates a horizontal component in the gradient of the index refraction, meaning the vertically-incident o-mode and x-mode waves will refract horizontally, with the o-mode diverging poleward and the

x-mode diverging equator-ward and the reflected wave of each wave returning along the same respective ray path [Rawer, 1993].

2.2 Power Reflected from the Ionosphere

Radar engineers are familiar with the radar range equation which expresses the radar's received power as a function of several variables:

$$P_r = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 L^2 R_t^2 R_r^2} \quad , \tag{2.24}$$

where S_r is the received power, P_t is the power output of the transmitter, G_t is the transmit antenna gain, G_r is the receive antenna gain, λ is the operating wavelength, σ is the reflection coefficient of the target, L is one-way path loss dissipation factor, R_t is the distance from the transmitter to target, and R_r is the distance from the receiver to target.

Ionospheric sounding is unique among radar applications in that HF waves incident upon the ionosphere are reflected specularly by a surface that is large compared to the antenna beamwidth. When all incident power is reflected by a large surface, the cross-section of the target increases proportionally to the square of the distance. In this special case the σ term offsets some of the losses in free-space propagation.

An alternative description of the monostatic radar system that is more relevant to

ionospheric sounding involves a transmit antenna distance *R* from the specular surface, and the receive antenna, a mirror image of the transmit antenna, at distance *R* behind the specular surface. The effective distance the signal travels is *2R* and the equation for received power reduces to a simple link budget equation between two points. At HF frequencies ohmic path losses are small except for waves of low frequency and at times of high D-region absorption. Ignoring these losses, the simplified expression for received power is, from the Friis Transmission Equation:

$$P_r = \frac{P_t G_t G_r}{(4\pi)^2} \frac{\lambda^2}{(2R)^2} = \frac{P_t G_t G_r}{64\pi^2} \frac{\lambda^2}{R^2} \quad . \tag{2.25}$$

The free-space spreading losses associated with the above expression are the most significant losses in the link budget. Example two-way path losses are illustrated below.

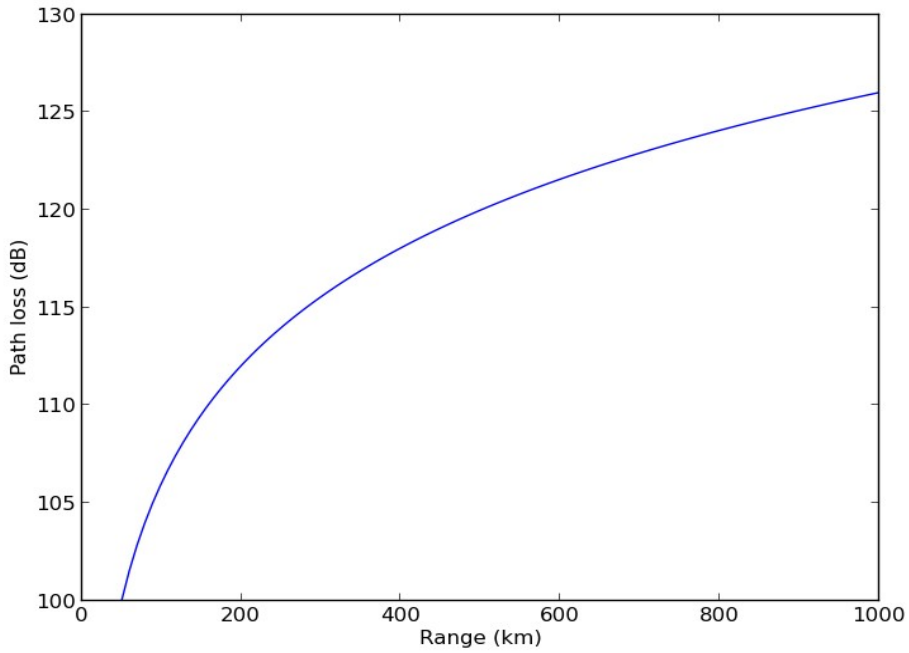

Figure 5: Two-way path loss vs. range to target layer

In practice there are additional factors not expressed in the above link budget equation. D-region absorption during daytime hours can cause significant signal attenuation. Another factor is variation in the ionospheric "surface". The ionospheric layer may not be perfectly horizontally stratified and instead be turbulent and/or have significant slant in the stratification of plasma density causing the direction of reflected power to be non-vertical. In the worst case little power is reflected back to the receiver and in the best case the ionospheric layers may have a concave shape and can magnify the incident wave, enhancing the received signal.

In general, the mid-latitude ionosphere exhibits the most ideal behavior in the

sense that low-altitude absorption is comparatively small and the ionosphere is largely horizontally stratified. The equatorial regions have less predictable stratification and significant low-altitude D-region absorption. Non-horizontal stratification of the ionosphere is also significant in the auroral regions, where particle precipitation can affect the local plasma structure, and in the polar regions, where solar radiation is relatively weak and vertical magnetic field lines allow for large mobility in the vertical direction. The above expression for received power is much simplified compared to the real world, but it is useful for generating a rough estimate of power at the receiver.

2.3 Noise in the HF Spectrum

A radio link consists of a transmitted signal passing through some channel to a receiver. The signal that is observed by the receiver is a channel-modified version of the transmitted signal plus noise. This noise obscures the signal of interest and is the fundamental obstacle in achieving a well-performing radio system. There may be many different sources of noise and they may be at different stages in the radio link. Noise may exist as electromagnetic wave energy and enter the receiver via the antenna, it may originate in the receiver due to thermal agitation of charge carriers in the components, or it may arise as an artificial product due to non-linearities in the receiver chain.

The fundamental lower limit for noise in HF radio is thermal noise, which is the random voltage fluctuations in the receiver that are caused by the Brownian motion of particles in the receiver circuit itself. The noise is spectrally flat ("white") and normally distributed in amplitude ("Gaussian") with its power P in watts a product of temperature T, bandwidth Δf, and Boltzmann's constant k_B:

$$P = k_B T \Delta f \quad , \tag{2.26}$$

or in terms of dBm at room temperature:

$$P_{dBm} = -174 + 10\log_{10}(\Delta f) \quad . \tag{2.27}$$

In HF applications there are often much more significant sources of noise than thermal noise, but for systems that employ much higher frequencies such as satellite links, microwave communications, or radio astronomy, thermal noise is a is the dominant noise source and can be minimized by reducing resistive components in the receiver, using low-noise pre-amplifiers that are optimally biased and impedance-matched, and in extreme cases by lowering the temperature of the receiver through refrigeration.

In shortwave applications most noise comes in the form of electromagnetic radiation that enters the receiver via the antenna. The amplitude of this noise is often expressed in reference to the thermal noise level, which would be the noise level if the system were completely isolated and in a perfectly quiet electromagnetic environment. For virtually any place on Earth, both natural and artificial processes create

electromagnetic noise in the MF and HF frequency bands, and these processes may be called atmospheric noise and artificial noise, respectively.

Atmospheric noise is caused by lightning. Dozens of powerful electrostatic discharges every second on Earth create wideband radio energy that propagates long distances. The amount of radio power generated globally by lightning is relatively constant with time, but the radio power at a particular receiver varies with location and time. Lightning occurs most often within 30 degrees latitude of the equator and atmospheric noise power is understandably greater in or near this region. The noise from this region propagates throughout the globe by skywave, but the ability for it to propagate depends on ionospheric conditions. A significant diurnal variation and a small seasonal variation exist due to D-region absorption, which attenuates the atmospheric noise. Therefore, atmospheric noise is generally greater at night.

Another significant source of natural noise on the HF band is galactic noise, caused especially by the sun and the greater Milky Way galaxy. However, this noise originates outside of Earth's ionosphere so it does not affect HF systems on the ground except at frequencies above the F-region peak critical frequency. Since the frequencies of interest in ionospheric sounding are generally below the peak critical frequency, galactic noise is mostly irrelevant to the concerns of ionospheric sounding.

Artificial noise comes from radio traffic on the numerous MF/HF bands and as

unintended radiation from home and industrial devices. Numerous regulations reduce the potential interference between competing radio users by designating usage bands, limiting users' transmit power and bandwidth, and suggesting rules of etiquette. Evidence of shortwave radio traffic can be detected everywhere on Earth, but quiet channels can be found depending on location, time of day, and frequency. Unintended electromagnetic radiation comes from street lights, electric motors, switches, and numerous other devices that may interfere with nearby shortwave receivers. Urban environments generally suffer much more from this unintended radiation than do rural environments.

The following figure illustrates the amount of noise above the thermal noise floor for an omni-directional antenna as a function of frequency.

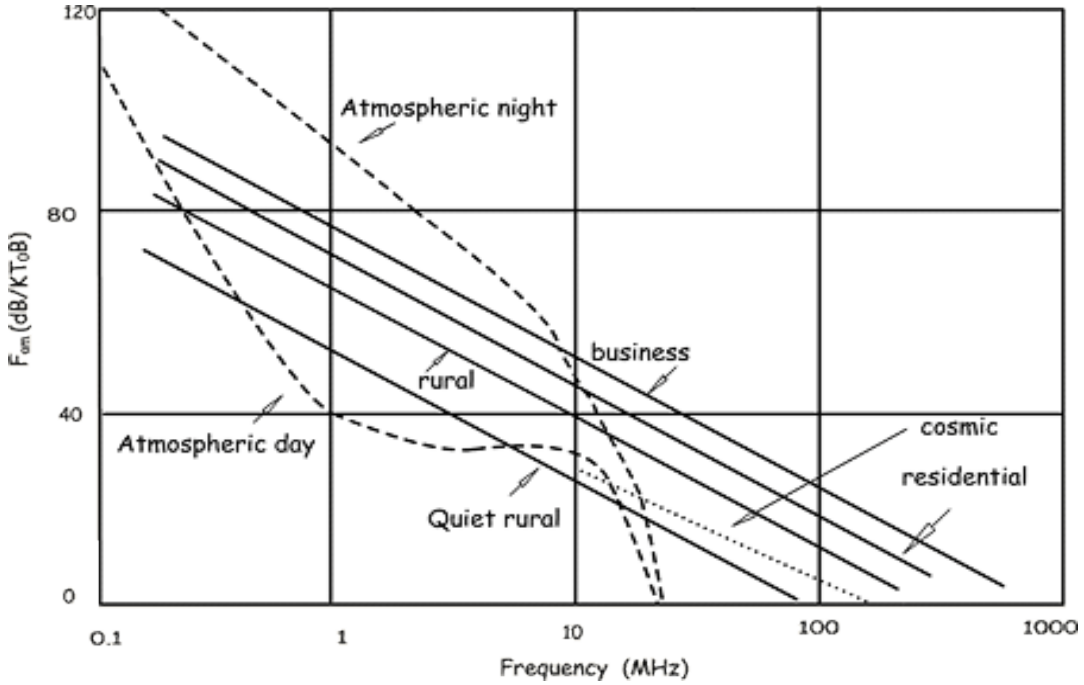

Figure 6: Typical atmospheric noise as a function of frequency and location

The range of operation for an ionosonde, roughly 1 to 20 MHz, contains sky noise that is 30 dB to as much as almost 80 dB above the thermal noise floor depending on the exact frequency, time, and location. The enormous amount of sky noise makes thermal noise a small and in most cases insignificant contributor to overall noise power. Methods to reduce sky noise include judicious choice of site location and good antenna directivity to null incoming noise sources.

The noise accepted by the antenna may of originate from different sources and have different amplitude distributions and spectral properties. Electromagnetic energy from lightening is has significant tails in its amplitude distribution, and is therefore

different from the normal distribution of thermal noise. Artificial noise often exists as narrow-band interference, making the background spectrum non-flat. It is incorrect to state that all noise in the HF spectrum is white and Gaussian, but for the next chapters that describe the signal processing techniques that reduce the effects of noise, it is mathematically convenient and a reasonable approximation to treat all noise as spectrally flat ("white") and normally distributed in amplitude ("Gaussian").

2.4 Ionograms

The basic data product of an ionosonde is an ionogram – a two-dimensional image representing the power reflected from the ionosphere as a function of frequency and height. The ionogram can be used to infer many useful parameters about the ionosphere, the most important being the critical frequency and height of the F layer peak. Other parameters include the density and height of each of the local ionospheric layers (D-, E-, F, etc.), total electron content in the atmosphere, and F-layer spreading. The process of converting an ionogram into specific parameters is known as ionogram reduction. This can be done by hand or automated by computer, which is more common and has become increasingly accurate over the years.

The details of ionogram reduction are described in the 1972 distribution

International Union of Radio Science (URSI) Handbook of Ionogram Interpretation and Reduction [Piggot & Rawer, 1972], which has standardized the techniques in ionogram reduction so that global ionospheric data can be compiled into a useful database. The URSI Handbook exceeds 300 pages and describes a surprising number of possible ionograms and their interpretations in excruciating detail. Thankfully, most parameters of the ionosphere under most conditions are simple to infer and do not require referencing specific case studies in the URSI Handbook. Computers have been used to programmatically reduce ionograms and have lessened the human workload immensely. For many of the world's operating ionosondes, the ionograms are automatically reduced and for some applications there is no need for them to be looked at with human eyes.

The following raw data is taken from the HAARP Digisonde in Gakona, AK on January 01, 2013 at 0010 UTC. Each cell is labeled numerically and is not yet scaled according to range and frequency. The cells are colored according to received power on a logarithmic scale, with red corresponding to highest power, blue to lowest, and yellow to intermediate.

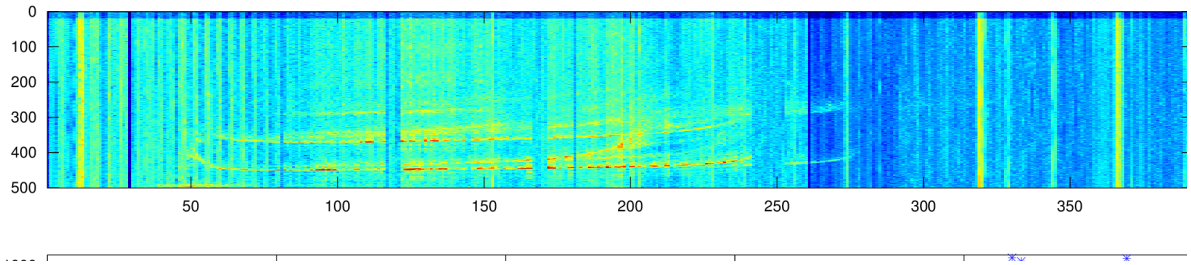

Figure 7: Raw ionosonde data from Gakona Digisonde, 01 January 2013, 0010 UTC

The highest-intensity powers correspond to the ionogram trace – the region in the ionosphere from which significant reflections are detected. This trace can be extracted by selecting only cells that exceed three standard deviations from the mean power level in each frequency column. The result is an ionogram similar to the figure shown below.

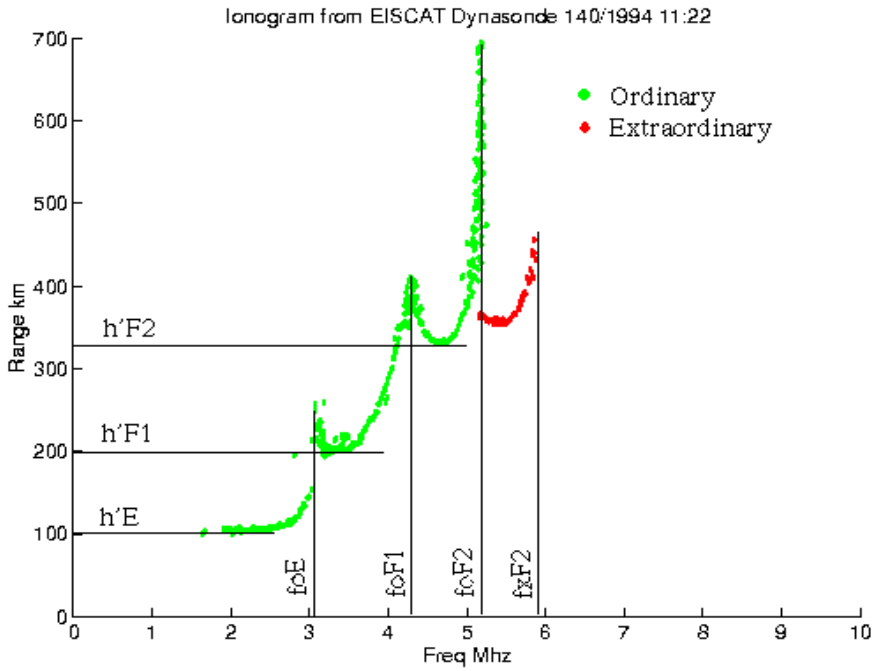

Figure 8: Example ionogram with clean traces

The vertical axis indicates the height of the reflection in kilometers, but does not compensate for the altered propagation of the signal in a plasma compared to free space. Thus, the vertical axis indicates virtual height rather than true height, which must be inferred in the process of ionogram reduction. The horizontal axis indicates sounding frequency in MHz.

The raw ionogram is composed of the first reflection of the ordinary mode and also several repeat echo traces. A second copy of the main trace occurs due to a "double hop" – the effect of a signal reflecting from the ionosphere, from the ground, and back

from the ionosphere one more time. All echoes except for the first are usually ignored in ionogram reduction except for the purpose of determining D-region absorption, where the reduction in signal power between successive echoes is an indication of attenuation in the atmosphere. Another extraneous echo trace is the x-mode trace, which is approximately a replica of the o-mode trace shifted by one-half of the plasma gyrofrequency. In the above ionogram the o-mode trace begins to turn steeply upward at approximately 5.2 MHz while the x-mode cusp doesn't occur until approximately 5.8 MHz. The o-mode trace generally yields most of the useful information about the ionosphere, and the o-mode trace is distinguished from the x-mode trace with the use of a polarized antenna.

The most immediately recognizable feature of the ionogram trace is the critical frequency. This is the cusp of the trace and a frequency above which there is no reflected power. The critical frequency of the o-mode is useful to radio operators because it indicates the highest usable frequency for near-vertical incidence skywave (NVIS) communication, and the critical frequency corresponds to the maximum ion density in the ionosphere according to the relation

$$N_e = \frac{\epsilon_0 m_e}{e^2} \omega^2 \quad , \tag{2.28}$$

where N_e is the electron density per cubic meter, ε_0 is the permittivity of free space, m_e is the mass of an electron in grams, e is the charge of an electron in coulombs, and ω is the

critical frequency of the plasma in radians per second. The above expression can be reduced to the simpler approximation:

$$N_e \simeq \left(\frac{f}{9}\right)^2 \quad , \tag{2.29}$$

where f is the critical frequency in Hz. The critical frequency can be used to find the maximum usable frequency (MUF), the highest frequency that will return to the ground by refraction in the ionosphere at arbitrary elevation angles. The MUF is given by:

$$MUF = f / \sin(\alpha) \quad , \tag{2.30}$$

where α is the elevation angle of the HF communication link.

In the above ionogram only one cusp is evident (the x-mode and o-mode cusps correspond to the same part of the ionosphere and are treated as a single cusp), but ionograms frequently exhibit multiple cusps below the f-region peak. These cusps correspond to the D, E, and F1 region peaks depending on their height, width, and time of occurrence. The frequency and range of each cusp yield the virtual height and local plasma density of the ionospheric layer.

Chapter 3 Radar Principles

Radar exploits the reflection of radio waves from a target to determine the distance to the target and/or the velocity of the target. A signal is transmitted from the radar toward the potential target, the radar listens for reflections from that target, and the round-trip time between transmission and reflection is a measure of the target's distance. Due to the Doppler effect, a measured change in frequency of the transmitted signal corresponds to a measure of the target's velocity relative to the radar.

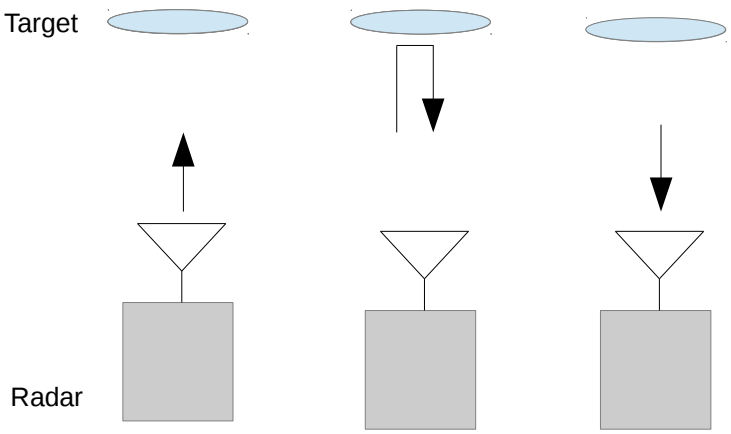

Figure 9: Radar transmission, reflection, and reception

3.1 Target Detection

The receive signal power is often very weak compared to the transmitted signal power

because the signal will travel a significant distance and perhaps only a fraction of the incident power will be reflected by the target. The challenge in radar design is being able to detect the possible reflection of a transmitted signal amongst noise. A higher signal to noise ratio (SNR) results in greater confidence in determining the existence of potential targets, as well as their position and velocity.

Since radars exist in an imperfect environment where noise corrupts the received signal, detection of a possible target is a matter of probabilities rather than certainty. The following figure illustrates a target that exceeds some threshold level in the presence of noise.

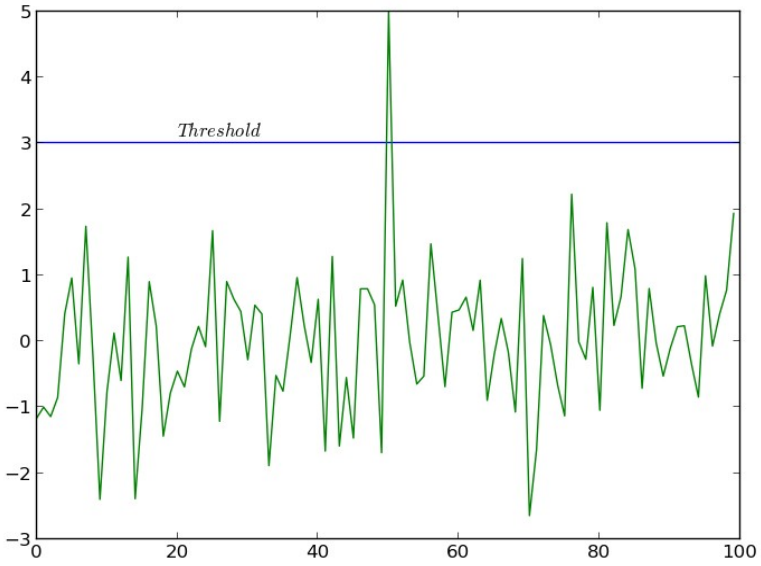

Figure 10: Target detection in the presence of noise

Due to the effects of noise, it is possible for a true target to go undetected because the noise subtracts from the signal amplitude, and it is possible for noise to appear as a target because the noise amplitude exceeds the threshold value. The likelihood of these scenarios can be expressed as a function of the signal amplitude, the expected noise amplitude, and the chosen threshold level.

Assuming white Gaussian noise, the magnitude of the noise amplitude follows a Rayleigh-distribution with first moment (noise amplitude) σ, while the signal plus noise amplitude is best described by a Rice distribution with zeroth moment (signal amplitude) ν, and first moment (noise amplitude) σ. The probability of a false alarm is the probability of the noise exceeding the threshold value, which is the area enclosed by the Rayleigh distribution and to the right of the threshold value. The probability of an undetected target is the area enclosed by the Rice distribution and to the left of the threshold value. The exact probabilities can be calculated by integrating each of the described areas, or they may be estimated using a Monte-Carlo simulation.

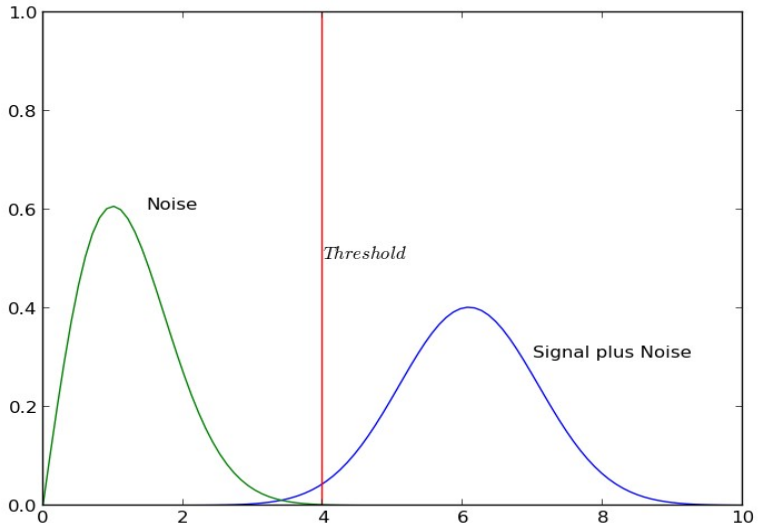

Figure 11: Probability distributions of signal and noise amplitudes

In choosing the threshold value there is a trade-off between minimizing the chance of false alarms and maximizing the chance of detecting true targets. Improving the signal-to-noise ratio helps achieve the best of these probabilities. The receiver operating curve (ROC) illustrates the relationship between probability of detection (True positive rate) and probability of false alarm (False positive rate) for a given SNR. It is desirable to choose a threshold level that is at the upper left on the ROC curve, which maximizes the true positive rate and minimizes the false positive rate.

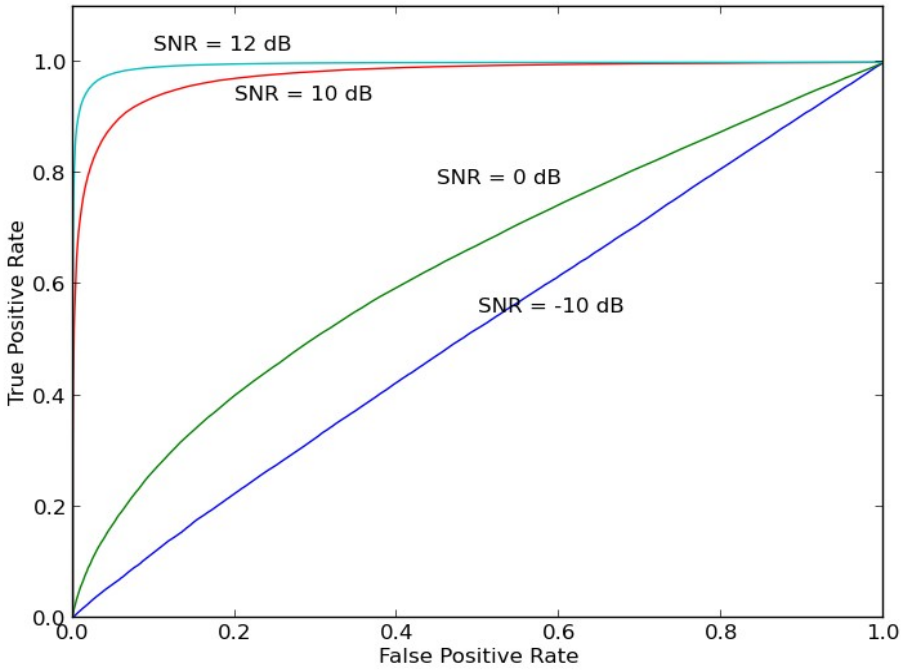

Figure 12: Radar operating curve

At the appropriate threshold level, 12 dB of SNR will give a 95% probability of detection while accepting only a 0.01% chance of false alarm. The quality of detection falls rapidly below 12 dB SNR, so any SNR greater than 12 dB will be considered a success in the design of the ionospheric sounder.

3.2 Range and Doppler Velocity

Pulsed-Doppler radar, which yields both range and velocity information of the target and offers the ability to improve the signal-to-noise ratio by integrating multiple

pulses, is most suitable for ionospheric sounding. A sequence of pulses is transmitted and the time t between the transmitted pulse and the received reflection determines the range R of the target according to:

$$R=\frac{ct}{2} \quad , \tag{3.1}$$

where c is the speed of the propagating signal, which is the speed of light in most cases. The factor of 2 accounts for the round-trip distance that the signal travels to the target and back.

If the target is moving, then the Doppler effect creates a frequency shift on the received signal carrier of wavelength λ, which is manifested as a phase shift $\Delta\theta$ between successive samples of the baseband signal. The phase shift between successive samples (pulses) separated by time T determines the Doppler velocity v of the target:

$$v=\frac{\lambda\Delta\theta}{2\pi T} \quad . \tag{3.2}$$

The measured Doppler velocity is a function of not only the phase shift, but the carrier wavelength λ. Shorter wavelengths will undergo a larger phase shift for a given target velocity, so higher frequencies are more sensitive to Doppler velocity.

3.3 Range-Doppler Ambiguity

Due to aliasing, the pulse repetition period *T* limits the maximum target range that can be unambiguously determined. For example, if a pulse is sent every 1 ms and reflections from a target occur 1.1 ms later, then the target's range is ambiguous – it appears the same as a target whose reflections occur at 0.1 ms after the pulses are sent. When choosing a pulse repetition period it is assumed that no significant power will be reflected from beyond the maximum unambiguous range, R_{max}:

$$R_{max} = \frac{cT}{2} \quad . \tag{3.3}$$

A similar limitation occurs in determining the Doppler velocity. If the phase shift between successive pulses is greater than 180 degrees, then the target's velocity is ambiguous. The choice of pulse repetition period determines the maximum unambiguous velocity, v_{max}:

$$v_{max} = \frac{\lambda}{2T} \quad . \tag{3.4}$$

Since the maximum unambiguous range is proportional to *T* and the maximum unambiguous velocity is inversely proportional to *T*, there is a trade-off known as range-Doppler ambiguity:

$$R_{max}\, v_{max} = \frac{c\lambda}{4} \quad . \tag{3.5}$$

Changing the pulse repetition frequency to increase the maximum unambiguous range simultaneously decreases the unambiguous Doppler velocity and vice versa.

It is possible to increase both unambiguous range and unambiguous velocity by employing non-uniform pulse periods, but in ionospheric sounding this is unnecessary. For example, if a 1000 km unambiguous range is desired with a 30 MHz (10 m) carrier, then the maximum unambiguous velocity is 750 m/s, an enormously large vertical velocity. Reasonable values for both unambiguous range and unambiguous velocity are easy to achieve even for the most extreme measurement ranges of the ionospheric sounding.

3.4 Resolution and Precision

The ability to measure target range may be expressed in range resolution, defined as the minimum distance step ΔR at which two targets of equal amplitude can be distinguished in range. It is related to the bandwidth BW of the transmitted pulse:

$$\Delta R = \frac{c}{2\,BW} \quad . \tag{3.6}$$

To increase the signal bandwidth the length of the rectangular pulse may be decreased and/or the rectangular pulse may be phase-coded to give higher bandwidth. Increased

signal bandwidth yields finer range resolution but comes at the cost of increased noise power in the received signal.

The ability to measure target range may also be defined in terms of precision, which may be the parameter more relevant to ionospheric sounding since there is only one target of interest. In a noiseless environment, the range precision is simply limited by the sampling frequency of the receiver ADC, but the addition of noise to the received signal places an uncertainty on detecting the time of receiving a reflected pulse from a target. The uncertainty in range may be expressed in terms of the uncertainty in measuring the threshold crossing of the received rectangular pulse. This discussion follows the geometrical argument given by [Levanon, 1988] and is illustrated below.

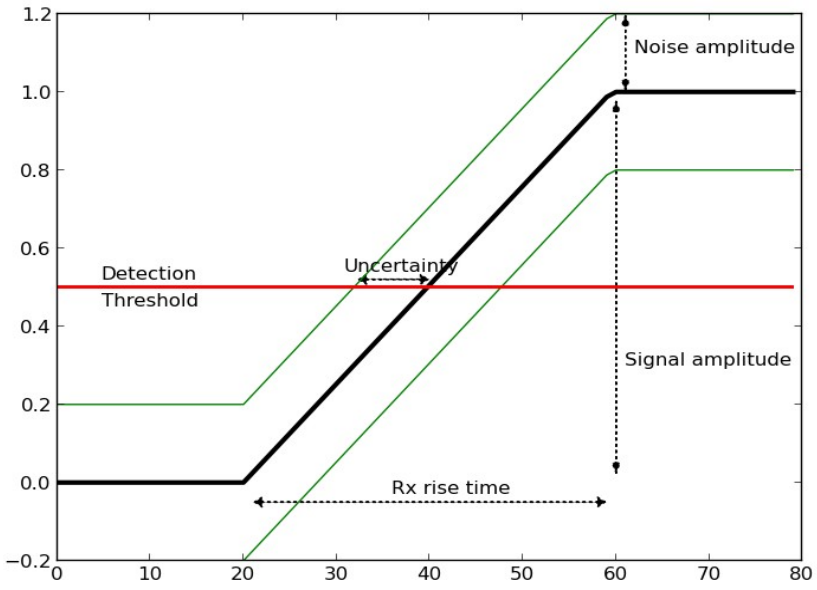

Figure 13: Range precision illustrated

The rise time of a filtered rectangular pulse t_R is determined by its bandwidth BW according to

$$t_R = 1/BW \quad . \tag{3.7}$$

The received signal consists of both signal amplitude s(t) and RMS noise amplitude n(t). The variance in the received signal amplitude created by the addition of noise imparts a variance in the time dimension. This can be expressed geometrically, where the triangle formed by the signal amplitude and rise time is similar to the triangle formed by the noise amplitude and uncertainty in time. Using the rules of similar triangles, the time of the target detection threshold crossing has uncertainty:

$$\Delta t = \frac{A\, n(t)}{t_R} \quad , \tag{3.8}$$

where A is the pulse amplitude, n(t) is RMS noise amplitude, and t_R is the rise time. The variance σ^2 in measuring the time of the threshold crossing is related to the pulse bandwidth BW (rise time) and the signal to noise ratio SNR:

$$\sigma^2 = \frac{1}{2\, BW\, SNR} \quad . \tag{3.9}$$

If both the rising and falling edges are used to determine the range, then the variance in the measurement falls by a factor of two, and converting the uncertainty in time to an uncertainty in range gives:

$$\sigma^2 = \frac{c}{8\,BW\,SNR} \quad . \tag{3.10}$$

The equation shows that increasing the signal-to-noise ratio and/or increasing the pulse bandwidth improves the range precision.

The ability to measure velocity may be qualified in terms of resolution, the minimum velocity step Δv at which two equal-amplitude targets may be distinguished in velocity and is related to the coherent integration time T_{CPI} of the transmission sequence:

$$\Delta v = \frac{\lambda}{2T_{CPI}} \quad . \tag{3.11}$$

Extending the integration time improves the velocity resolution at the expense of greater transmitted energy and longer coherency time. Velocity resolution is important in certain radar applications where a potential moving target must be distinguished from stationary ground clutter, but for ionospheric sounding there is no ground clutter and it is generally useful only to know if the target is closing (downward moving ionosphere) or opening (upward moving ionosphere), and so fine velocity resolution is unimportant.

The precision of the velocity measurement is similar to the precision in range measurement. [Richards et al., 2010] shows that the variance σ_f^2 in the Doppler frequency measurement is a function of the sampling frequency (pulse-repetition frequency) f_s, the received signal-to-noise ratio *SNR*, and the number of samples *M* in the time-series:

$$\sigma_f^2 \simeq \frac{3 f_s^2}{\pi^2 SNR \cdot M^3} \quad . \tag{3.12}$$

Again, knowing the exact vertical velocity of the ionosphere is not critical to most studies, so no extraordinary efforts are made in the signal processing to extract high precision in velocity. However, good precision is an unintended consequence of long integration periods used for improving the signal-to-noise ratio.

3.5 Multi-pulse Integration

The idea of multi-pulse integration is common to all data-collection methods, where variation in individual measurements (noise) is reduced by collecting many samples of data and and integrating them (averaging). There are different ways to implement this concept for radar signal processing, especially with regard to whether signals are integrated coherently (phase information is preserved), or non-coherently (phase unknown).

Non-coherent integration may be done by summing the power levels of multiple bins of the same range. The integration is non-coherent, because the power level is a positive scalar and contains no phase information. Mathematically, the amplitude *a* in each individual range bin may be described as the sum of the received radar signal *s* and

Gaussian noise n with zero mean and variance σ^2:

$$a = s + n(0, \sigma^2) \quad . \tag{3.13}$$

The power P of the received signal is the squared magnitude of the above quantity, where the middle term is zero when the signal and noise are uncorrelated. The Gaussian distribution of noise amplitude becomes a chi-square distribution of noise power with mean σ^2 and variance $2\sigma^2$:

$$P = s^2 + \chi(\sigma^2, 2\sigma^2) \quad . \tag{3.14}$$

Summing the power from multiple range bins results in an increase in the signal power by a factor of N, while the variance in the chi-square distribution increases by a factor of $N^{1/2}$. Therefore, the net signal-to-noise ratio increases by $N^{1/2}$.

Coherent integration is done by summing the amplitude levels of multiple bins in the same range. The amplitude in each range bin is the sum of the signal s and the Gaussian noise variable n with zero mean and variance σ^2:

$$a = s + n(0, \sigma^2) \quad . \tag{3.15}$$

N signal amplitudes are added. The signal amplitude increases by a factor of N, and the variance of the normally-distributed noise increases by a factor of $N^{1/2}$. Squaring the amplitude gives:

$$P = N^2 s^2 + \chi(N\sigma^2, 2N\sigma^2) \quad . \tag{3.16}$$

Coherent integration yields a signal power improvement of N^2 and a noise increase of N,

for a net SNR improvement of *N*.

The signal processing of the Doppler time series may also be viewed in terms of a transformation from the time-domain to the frequency domain. The signal of interest from *N* successive pulses is a complex time series with amplitude *A* and some frequency ω:

$$s[n] = Ae^{-j\frac{\omega \cdot n}{2\pi}} \quad , \tag{3.17}$$

whose total energy is

$$E_s = \sum_{i=0}^{N} |s[n]|^2 = N \cdot A^2 \quad . \tag{3.18}$$

The noise is a complex random variable whose two components are normally-distributed with mean zero and expected amplitude *σ/sqrt(2):*

$$n_{norm}[n] = Normal\left(0, \frac{\sigma^2}{2}\right) + j \cdot Normal\left(0, \frac{\sigma^2}{2}\right) \quad , \tag{3.19}$$

whose expected total energy is

$$E_n = \sum_{0}^{N} |n[n]|^2 = N \cdot \sigma^2 \quad . \tag{3.20}$$

The SNR in the time series before Doppler processing, as expected, is

$$SNR = \frac{P_s}{P_n} = \frac{N \cdot A^2}{N \cdot \sigma^2} = \frac{A^2}{\sigma^2} \quad . \tag{3.21}$$

Taking the discrete Fourier transform of the *N*-length time series transforms the signal to the frequency domain:

$$S[k]=\sum_{n=0}^{N-1} s[n]e^{\frac{-j2\pi kn}{N}} \quad , \tag{3.22}$$

If the time series contains a single frequency component $\omega k/2\pi$ that falls on some point k in the DFT, as would be the case for a stationary radar target, the DFT simplifies to:

$$S[k]=N\cdot A\cdot\delta[k-\frac{\omega k}{2\pi}] \quad , \tag{3.23}$$

where the total signal energy is again

$$E_s=\frac{1}{N}\sum_{k=0}^{N-1}|S[k]|^2=N\cdot A^2 \quad . \tag{3.24}$$

The normally-distributed series of noise amplitudes in the time-domain transforms to a normally-distributed series of amplitudes in the frequency-domain, with an amplitude increase of $N^{1/2}$:

$$N_{norm}[k]=\sum_{n=0}^{N-1} n_{norm}[n]e^{\frac{-j2\pi kn}{N}}=\sqrt{N}\cdot n_{norm}[k] \quad , \tag{3.25}$$

and the total noise energy is again

$$E_n=\frac{1}{N}\sum_{k=0}^{N-1}|N_{norm}[k]|^2=\frac{1}{N}\sum_{n=0}^{N-1}|(\sqrt{N}\cdot n_{norm}[n])|^2=N\cdot\sigma^2 \quad . \tag{3.26}$$

Parseval's theorem, which states that the energy of the time-domain expression of a signal is equal to the energy of the frequency-domain expression, is conserved for both signal and noise energies. There is no improvement in the signal-to-noise ratio of the above signals until we consider the fact that the signal falls in a single spectral bin while the

noise energy is uniformly distributed. For the spectral bin of interest, the signal energy remains the same, while the noise energy falls by a factor of *N*.

$$SNR = \frac{E_s}{E_n} = \frac{N \cdot A^2}{\sigma^2} \quad . \tag{3.27}$$

As shown previously, the SNR improves by a factor of *N* after coherent Doppler-processing.

Coherent integration is the champion method for coding gain, and comes with the added benefit of preserving phase information and thus the possibility of extracting the Doppler velocity information from the received signal. When it is possible to construct a phase-coherent transmitter-receiver system it is the preferred method of multi-pulse integration.

Coherent integration requires that the target have the same velocity during the entire coherent processing interval, but for ionospheric sounding the target can remain coherent for on the order of 10's of seconds [Reinisch, 2007], so hundreds or even thousands of pulses can be transmitted, received, and then meaningfully integrated.

3.6 Pulse Compression

Intra-pulse modulation of a transmitted signal can be used to increase the bandwidth of a signal without sacrificing the pulse energy. Alternatively, it extends the

length of a pulse and hence its energy without sacrificing bandwidth (range resolution). The added pulse energy can be harvested to yield a higher signal-to-noise ratio by correlating the received signal with the known modulation pattern. A pseudo-random modulation pattern with an autocorrelation function that has a peak at lag zero and small for all other lags is desired.

The argument for coding gain is similar to that of multi-pulse integration. Cross-correlating the received signal with the N-length transmit phase code yields a factor of N improvement in signal amplitude, while cross-correlating the received noise with the pseudo-random transmit phase code (the two are uncorrelated) yields a $N^{1/2}$ increase in noise amplitude. The signal power increases with N^2 and the noise power increases with N for a net SNR improvement of N.

There are many pseudo-random sequences that are viable for radar. The choice of which depends on several factors: desired gain, side-lobe amplitudes, and Doppler tolerance. In general, a longer code length is desirable because it gives a larger SNR gain. However, there are costs to this. The first is the blind regions that are imposed by a long transmission sequence. The radar is unable to receive any potential reflections during the time it is transmitting, so the length of the pulse sequence limits the first viewable range of the radar. Also, since the matched filter only produces the full gain when it receives the complete length of the pulse code, the effective blind region is twice

the length of the transmitted pulse. A second possible drawback to long pulse-codes is the side-lobe amplitudes of the autocorrelation function of the pulse sequence. The Barker codes are the only known binary code sequences that have autocorrelation side-lobe amplitudes less than or equal to the reciprocal of the code length, and the longest Barker code length is 13. Longer codes require imperfect codes and thus allowing for larger side-lobe amplitudes in the autocorrelation function, which increases the probability of false detection and/or decreases the ability to resolve two distinct targets. The final cost of long pulse codes is decreased Doppler tolerance. Matched filtering at the receiver assumes that the received signal is a time-shifted, amplitude-scaled version of the transmitted pulse. However, the shape of the received signal pulse is altered according to the amount of Doppler-shift, and so the strength of the matched filter function is decreased. In general, longer pulse sequences are less tolerant to Doppler shift than are short pulse sequences since the received signal has to be correlated over the length of the pulse.

In ionospheric sounding there is only one target of interest, probability of false alarm is expected to be low, and the amount of Doppler shift is relatively small. Therefore, the largest factor in deciding on what pulse code to use is allowable blind ranges in the received signal grid. The first range of interest is never less than 50 km, and often greater than 100 km, depending on the frequency. Therefore, the length of the

pulse sequence may be on the order of 300 μs. For a bandwidth of 25 kHz (50 symbols/second) the number of chips in the pulse sequence may be on the order of 15, giving a first viewable range of 50 km. A length-30 pulse code gives a first viewable range of 100 km.

3.7 Practical Limits of Performance

Coding gains offer much improvement in SNR and have allowed ionosondes in the past few decades to move from multi-kilowatt transmitters to several hundred watt transmitters. As a result, ionosondes have become cheaper, smaller, and more reliable. However, it is important to discuss the practical limits involved in increasing coding gain to make up for decreasing transmit power.

The combined process of pulse integration and matched filtering is a linear operation in the formal sense that

$$F(a \cdot x) = a \cdot F(x) \quad , \tag{3.28}$$

and

$$F(x+y) = F(x) + F(y) \quad . \tag{3.29}$$

Informally, this means that some modification of the input x yields the same modification of the output $F(x)$. An input signal of increased amplitude yields a corresponding

processed signal of increased amplitude. An input signal plus noise yields a processed signal plus processed noise. The upshot is that even the most minute signal can be brought to usable SNR levels if the processing gain of the function *F(x)* is great enough.

Assuming linearity of the entire system -- the transmitter, radio channel, receiver, and processing chain-- in theory it would be possible to use a transmitter of vanishingly small output power and then make up for it in coding gain by increasing the transmit pulse length and integrating an increasingly large number of pulses to achieve the desired SNR. Even while ignoring the fact that the ionosphere will not remain a coherent target for increasingly large integration times, a realistic a system is limited by non-linear effects, especially in the receiver chain. The dynamic range of the ADC limits the linear range of the receiver, because at extremely low signal levels the ADC is unable to resolve very small voltages. The receive pre-amplifiers may also impose a limit on the linearity of the receiver in the form of third harmonic distortion. The radar link budget must satisfy some minimum signal level at the receiver to achieve linearity, only at which point is it possible to tease a small signal out of the noise by signal processing techniques.

This thesis aims to address the potential demand for low-energy, portable ionospheric sounders. For many radar systems the energy costs are a secondary concern, but for ionospheric sounding in remote areas it can be paramount. The above sections analyzed the link budget of the proposed ionospheric sounder and the signal-to-noise

ratios of the raw and processed signals. Another relevant performance metric is signal energy efficiency. This last metric is necessary because signal processing enhancements such as coherent integration and coded pulses offer reduced transmit power, but come at the expense of a proportional increase in transmit time and increased computational cost. *Coding gain offers no fundamental improvement in energy efficiency.* In practice there may be a significantly-improved DC to RF conversion efficiency offered by using a large coding gain and correspondingly lower-power transmitter but the theoretical limits to energy efficiency are unchanged by signal processing. The issue of energy efficiency must instead be addressed by way of analog component design, antenna selection, site selection, and sounding scheduling.

Chapter 4 Survey of Current Systems

It would be foolish to design a new ionospheric sounder without an eye on existing systems, and it would be unfair to benefit from the current knowledge base without crediting its sources. The previous chapter discussed principles that are shared by many radar systems and are common knowledge among radar engineers. The current chapter highlights things specific to ionospheric sounding and provides a bridge between the potentially naive application of theory and the practical results gained by experimentation. The ionospheric sounding systems that provide the material of this section are the Digisonde by University of Massachusetts Lowell Center for Atmospheric Research [Reinisch, 2007] and the Vertical Incidence Pulsed Ionospheric Radar (VIPIR) unit by Scion Associates [Bullet, 2008]. The first UMLCAR Digisonde system was created in 1970's and its latest generation of ionsondes are somewhat of an industry standard in the United States. The first VIPIR system was put into service in 2008 at the University of Colorado, Boulder ionospheric sounding station and seeks to be the next generation in ionospheric sounding systems.

4.1 Coherent Transmission/Reception and Digital Systems

The earliest HF transmitters employed tube amplifiers and analog circuitry that were relatively crude compared to the solid state amplifiers, crystal oscillators, and digital systems of today. Advances in hardware have allowed radio systems to operate coherently in phase and perform signal processing operations in the digital domain. The trend in radio design is for all signal processing operations to be done digitally, sometimes by application-specific integrated circuits, but often in software on a general-purpose computer, so that the digital system is as "close" to the antenna as possible. Both the Digisonde and VIPIR systems employ digital hardware and coherent systems that allow the ionosonde to achieve large coding gains and reduce the necessary transmitter power, but their approaches to conversion between the digital and analog domains are different.

The Digisonde uses a dedicated unit each for GPS-disciplined clock generation, signal timing, RF oscillator generation, waveform generation, receiver mixer, and digitizer. The Digisonde uses numerous clock signals to perform soundings, mainly because there are several IF frequencies and mixing stages in both the transmit and receive paths. Coherence between the transmit and receive paths is achieved since the transmit and receive mixer frequencies generated by the RF oscillator unit are phase-

locked.

The VIPIR system has benefited from recent advances in the speed of ADC's and DAC's and uses a simpler approach to digitization and phase coherency, where a single clock provides a reference to the direct-conversion ADC receiver and DAC synthesizer. The VIPIR uses separate units for waveform generation and RF sampling that share the same 80 MHz clock reference and are connected to the host computer on the same USB bus. The VIPIR is the much simpler unit in terms of hardware, with fewer components and a much greater share of the signal-processing burden accomplished digitally.

4.2 Phase-Coded Pulses

Phase-coded pulses (intra-pulse modulation, pulse compression) improve the signal-to-noise ratio of the received radar pulse by extending the length of the transmitted pulse, effectively increasing the amount of transmitted energy but without sacrificing bandwidth and range resolution. There are numerous phase modulation schemes that attempt to achieve large zero-lag autocorrelation and small sidelobes.

The Barker codes are the only known optimum binary sequences in that the autocorrelation sidelobes are of uniform amplitude and are therefore popular in many radar and telecommunication applications. The longest Barker code is 13-chips, so

longer pulses require other types of pseudo-random sequences such as complementary coded sequences, maximum length sequences, and Kasami sequences. The choice of which code depends on the desired coding gain, the magnitude of the side-lobes of the autocorrelation function, and the strength of the matched filter response in the presence of a Doppler-shift of the received signal.

The makers of the Digisonde were very concerned with the side-lobe amplitudes of the pulse-sequence autocorrelation function, and employ complementary-code sequence pairs. The autocorrelation function side-lobes of the first pulse code are opposite phase to the autocorrelation function side-lobes of the second pulse code. By transmitting a receiving each pulse code alternately and summing the two, the side-lobes effectively cancel and the complementary code sidelobes are perfectly zero.

The VIPIR also uses a coded pulse, but the type is unknown to the author at the time of this writing.

4.3 Coherent Integration of Multiple Pulses

In any research field it is good and often necessary to take an average of many samples because it reduces the contribution of noise that is present in each individual measurement. In radar, a simple average of received power in each range bin for N

pulses (non-coherent integration) yields a noise power increase of $N^{1/2}$ and a signal power increase of N for a net SNR improvement of $N^{1/2}$. When phase information can be preserved, a coherent integration is preferred as it yields a noise power increase of N and a signal power increase of N^2 for a net SNR improvement of N.

Both the UMLCAR Digisonde and the VIPIR instruments are phase-coherent systems and perform coherent integrations of multiple pulses. The UMLCAR Digisonde employs a 128-pulse set for a coding gain SNR improvement of 21 dB, and the VIPIR instrument employs an 8-pulse set for an improvement of 9 dB. The two instruments perform similar processing routines to achieve some amount of coding gain, although the Digisonde is more aggressive in terms of the number of integrations.

4.4 Phased Antenna Arrays

When discussing phased antenna arrays, the design of the antenna hardware and the design of the radar signal processing algorithms start to overlap. The use of a phased antenna array may be viewed as either a single high-gain antenna that is electronically steerable or as multiple antennas from which a single, enhanced signal can be created algorithmically. In either case, the use of phased antennas increases the amount of signal power to/from the desired beam direction by a factor proportional to the square of the

number of antennas in the array.

In a simple ionosonde configuration, the transmit and receive antennas are the same, and a switch determines whether the antenna is in transmit mode or receive mode. However, both the Digisonde and VIPIR instruments use a single large antenna for transmitting, and an array of small antennas for receiving. This approach has become somewhat of a standard in ionospheric sounding, due to the improved directivity of the receive antenna array and the ability to determine the angle of arrival of the received signal through interferometry techniques.

The choice of transmit antenna is largely determined by site and budget restrictions, since the long wavelengths of ionospheric sounding demand large towers to support the wire of the transmit antenna. The VIPIR group favors a single inverted delta transmit antenna. The Digisonde group favors a single antenna or two antennas in a crossed configuration of either the vertical-rhombic or the delta type. In any case, large vertical gain across the entire ionosonde frequency range is the desirable quality, with the Digisonde group also desiring the ability to selectively transmit linear- or circularly-polarized waves.

For the receive antennas, both the Digisonde and the VIPIR employ an array of four small receive antennas. Unlike the transmit antenna, the efficiency and thus the size of the antenna are not critical. Using a low efficiency antenna lowers the amount of

signal received, but reduces the amount of HF atmospheric noise in equal proportion, so the received SNR is not significantly deteriorated. The signal-to-noise ratio can be improved, however, by increasing the antenna directivity, which decreases the amount of received noise relative to the signal power. This improvement in SNR can be achieved as long as the radar target is small compared to the antenna beam-width. The antenna array layout is the same for both instruments, with three antennas forming a triangle and a fourth antenna in the center. Both instruments sample each of the four signals digitally, so angle-of-arrival can be determined algorithmically by processing and combining the four channels digitally. The Digisonde calculates the signal power in seven discrete azimuth/elevation beam directions by applying the appropriate phase shift to each channel and summing them, then deduces angle-of-arrival by finding which of the seven beam direction contains the largest signal power. In contrast, the VIPIR instrument searches for the largest power azimuth/elevation beam angle by cross-correlating the four receive channels as a function of phase shift. Rather than seven discrete beam angles the VIPIR effectively calculates a continuous range of beam angles to determine maximum power in the angle-of-arrival.

4.5 O- and X-mode Discrimination

The ordinary and extraordinary modes of propagation in the ionosphere have a different dispersion relation and opposite polarizations – the o-mode being left-hand circularly polarized at the receive antenna in the northern hemisphere and the x-mode being right-hand circularly polarized. It is possible to discriminate between the two modes with certainty by using two receive antennas in a crossed configuration. A circularly polarized signal will be received by both antennas simultaneously, but the phase of one antenna will either lead or lag the other antenna by 90 degrees depending if the polarization is left- or right-hand polarized. Occasionally the discrimination is complicated by the existence of a z-mode echo that has the same polarization as the o-mode. The z-mode then must be discriminated by other means.

O- and x-mode discrimination is common practice in modern ionosondes, but the implementation varies. The Digisonde combines the two antenna signals in the analog domain using a 90 degree phase-shifter/combiner. Receiving the left- or right-hand circular polarization is selected prior to each processing interval with an electronic switch, so acquiring both wave polarizations requires two separate processing intervals. In contrast, the VIPIR instruments sample both polarizations for each of the four receive antennas, producing a total of eight receive channels. The o- and x-modes are produced by digitally phase-shifting and combining the linearly-polarized pairs.

Chapter 5 System Description

The goal of this ionosonde design is to produce an instrument capable of producing ionograms with reduced size, cost, and power consumption compared to existing systems. Basic knowledge of the ionosphere (Chapter 2) and basic radar principles (Chapter 3) provide a starting point in system design, and a review of existing systems (Chapter 4) add practical knowledge of what is possible. The current chapter describes the design of the system.

5.1 Design Approach

Two objectives guide the design of the ionospheric sounder. The first objective is to exploit the advantages of modern digital hardware and open-source software, which allows the performance goals to be achieved with minimal development and maintenance costs. The second objective is to design for flexibility in the sounder, which would allow the system to follow dynamic sounding schedules, compensate for potential shortcomings of the analog components, and be adaptable to future experimental modes of operation.

The ionospheric sounder can be broken down into four physical subsystems: the host computer, the Ettus Research Universal Software-defined Radio Peripheral (USRP),

the analog front-end, and the antenna (see figure below).

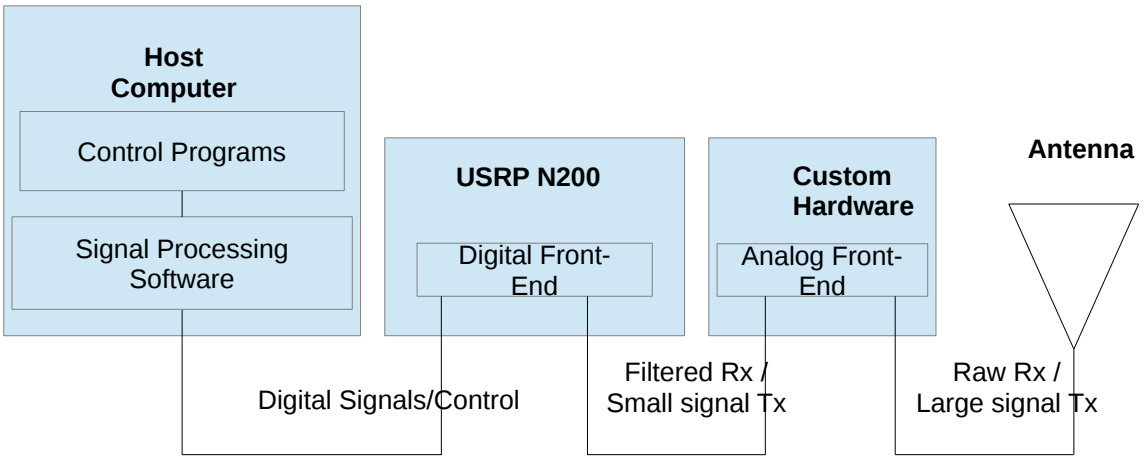

Figure 14: Ionosonde functional block diagram

In general, the performance quality of the analog front-end and antenna are somewhat uncertain, being determined by site and budget limitations. However, poor performance can often be overcome by increasing the number and length of pulses transmitted by the radar. Knowing this, the design focus is on the computer and USRP, which are not variable with budget or site considerations and, if properly designed, can easily compensate for poor performance in the other system components.

The functionalities of the computer and USRP can be broken down into the following pieces: control, signal processing, and the digital front-end.

The computer is host to the control programs and the signal processing software. The control programs handle the high-level data control like reading and writing data files, initiating sounding operations, uploading results to a computer network, etc. The

signal processing software performs waveform generation, filtering of received signals, and Doppler processing of the radar returns. Since both systems are housed on a general purpose computer and implemented with software, their functions may be altered quickly and inexpensively. All operations are repeatable and inexpensive to perform on the low bandwidth digital signals.

The USRP houses the digital front-end, which performs conversion between the relatively low bandwidth digital signals of the computer and the high frequency analog signals of radio waves. It changes the sample rate using a high-speed FPGA, changes the center frequency of digital signals with numerical oscillators and digital mixers, and converts between the analog and digital domains by using high speed ADC's and DAC's. Since the USRP samples the HF spectrum directly and performs mixing and down-sampling in the digital domain, the behavior of its operations match perfectly with the expected behavior, and they are repeatable. The USRP interfaces conveniently with a general-purpose computer, so its functions can be controlled quickly and easily through the host computer.

5.2 Overview of the Ettus Research USRP

Ettus Research offers a line of Universal Software-Defined Radios (USRP's) that

serve as the connection between the digital domain on a host computer and the analog domain of radio signals. The USRP model name describes the host interface and generation, where the USRP N200 used in this project is described by N for networked model, 2xx for second generation, and the 00 distinguishes the N200 from the N210, which has a larger FPGA. The hardware and software are well-supported through the on-line knowledge base managed by Ettus Research, and basic operation is described in the USRP-N2xx device manual [Ettus Research, 2014].

A USRP device is tailored to a particular RF band through the use of a specific daughterboard, an analog front-end that may, depending on the model, amplify and/or mix the RF signal down to an intermediate frequency. For this project, the USRP was used in conjunction with the LFRX and LFTX daughterboards which perform no mixing but have unity gain buffers and 30 MHz low-pass filters, allowing the USRP to operate directly on the RF spectrum from DC to 30 MHz.

The USRP N200 samples the RF spectrum with two 100 MHz, 12-bit ADC's and generates RF with two 400 MHz, 14-bit DAC's.

In addition to the sampling and synthesis functions provided by the ADC's and DAC's, the primary functions of the USRP are digital down-conversion (DDC) and digital up-conversion (DUC), which allow the band-limited RF signal of interest to be represented by a sample rate that is manageable by the host computer. The digital output

of the two receive channels A and B may be routed to the DDC in one of several ways, depending on the user's application and whether the RF signal is mixed to an intermediate frequency before sampling or sampled directly . If the ADC's are sampling two components of a complex signal at an intermediate frequency, then A can sample the in-phase component (I) and B can sample the quadrature-phase component (Q), or vice-versa (B samples I and A samples Q). The two components of the complex signal are then passed to the DDC chain. If the ADC's are sampling the RF spectrum directly, then one or both of the channels can operate in “real-mode”, meaning that when the signal is passed to the mixer the Q component of the signal is assigned sample values of zero. The USRP contains two complete DDC chains that may allow independent processing of two real signals.

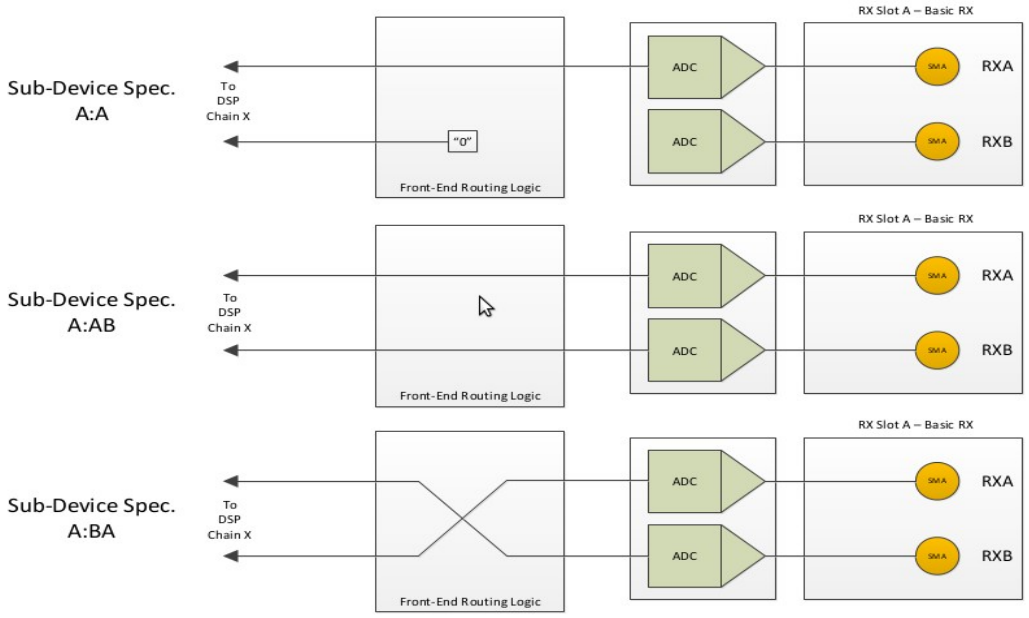

Figure 15: USRP signal routing

The DUC is similar to the DDC, in that it can output either a single real-mode signal or two components of a complex-mode signal. But unlike the receiver which has two DDC chains, there is only one DUC chain and so the USRP is unable to transmit two independent real-mode signals. The two DAC outputs are tied together in a single DUC chain.

The DDC chains and DUC chain exist in the USRP's FPGA fabric, and consist of numerically-controlled oscillators (NCO), filters, and decimators/interpolators. Their behavior is modified in terms of their NCO frequencies, filter bandwidths, and decimation/interpolation rates by writing certain values to the FPGA registers. The user requests a desired RF center frequency and host-side sample rate, and the host computer,

via the hardware driver, writes the appropriate register values to the USRP's FPGA. The DDC chain block diagram is illustrated below.

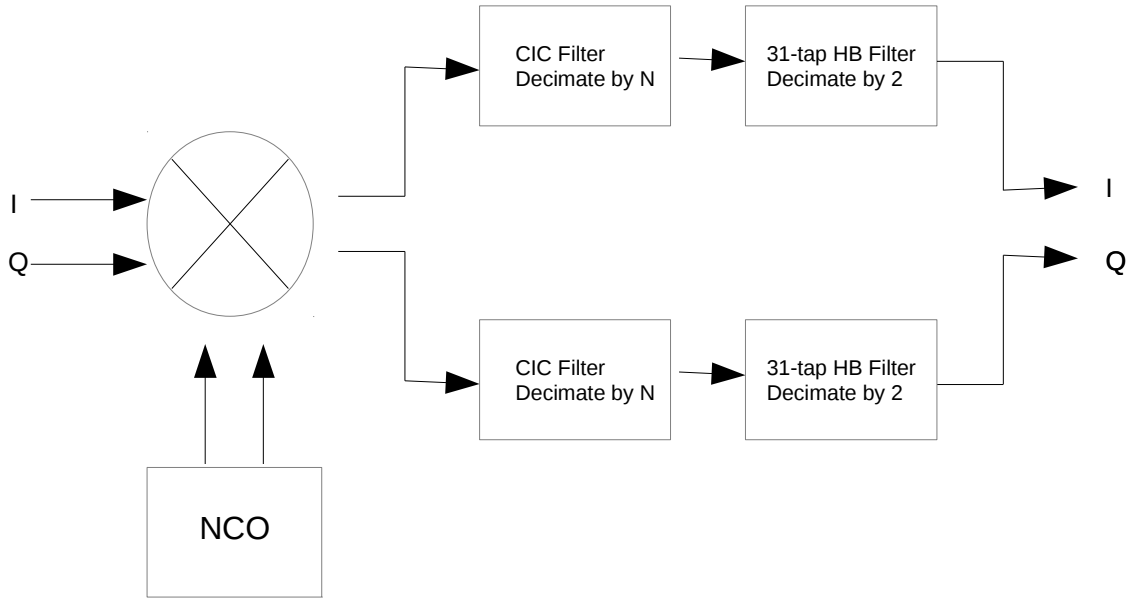

Figure 16: USRP DDC chain

The USRP N200 communicates with the host computer via a 1 Gbit Ethernet link, which limits the effective bandwidth of the radio signals that can be transferred. On receive, the signal format is either complex pairs of 16-bit signed integers or complex pairs of 8-bit signed shorts. Thus the theoretical upper bound on bandwidth when the device is only receiving is 31.25 MHz or 62.5 MHz for 16-bit integer format and 8-bit short format, respectively. The actual maximum bandwidth is less due to overhead in the packet structure, and the advertised values for maximum bandwidth are 25 MHz and 50 MHz for 16-bit integer format and 8-bit short format, respectively. When the device is

transmitting and receiving simultaneously, the Ethernet bandwidth must be split between the transmit and receive streams. In ionospheric sounding the bandwidth of a single sounding channel is less than 100 kHz and so the capacity of the Ethernet link is more than adequate.

The software interface between host computer and USRP is the Universal Hardware Driver (UHD) which is thoroughly documented on the web [Ettus Research, 2013]. The UHD is written in C++ and is an object-oriented Application Programming Interface (API) that provides convenient mechanisms for passing data to and from the device, creating data structures, and setting control parameters on the USRP. Creating transmit streams or receive streams for the USRP translates to creating instances of API-defined classes. The USRP hardware is configured by initializing a *multi_usrp* object and then using its public functions to set the clock reference, output center frequency, sampling rate, etc. A receive stream is created by initializing a *uhd::rx_streamer* object, and the stream begins execution by calling the *uhd::rx_streamer* member function *recv()*. Data is returned to the user via a pointer to a vector of received samples. Similarly, a transmit stream is created by initializing a *uhd::tx_streamer* and the stream begins execution by calling the *uhd::tx_streamer* member function *send()*, which points to a vector of samples that are to be transmitted.

5.3 Using the USRP as a Radar

Radar is like any radio application in that information must be sent and received over some RF channel, but unlike most radio applications the timing between transmitted and received streams is critical – the precise time and phase of the received signal in relation to the transmitted signal must be known. This can be accomplished in one of two ways: (1) The receive stream begins before the transmit stream, some part of the transmit signal is coupled back to the receive path, and post-analysis of the received data reveals exactly when transmission occurred. (2) The transmit and receive streams begin simultaneously (or at some precise time difference) using some internal or external trigger.

Method 2 is preferred here, since it is simpler and made possible by the high-resolution internal triggering in the second generation of USRP's. Command arguments to the receive and transmit streams yield timing resolution of 10 ns, corresponding to the USRP's 100 MHz reference clock. The time critical functions are implemented using the UHD and the C++ API in a program called *sounder*. This program handles the basic functionality of sending and receiving data to and from the USRP. The following figure illustrates the flow control of *sounder*.

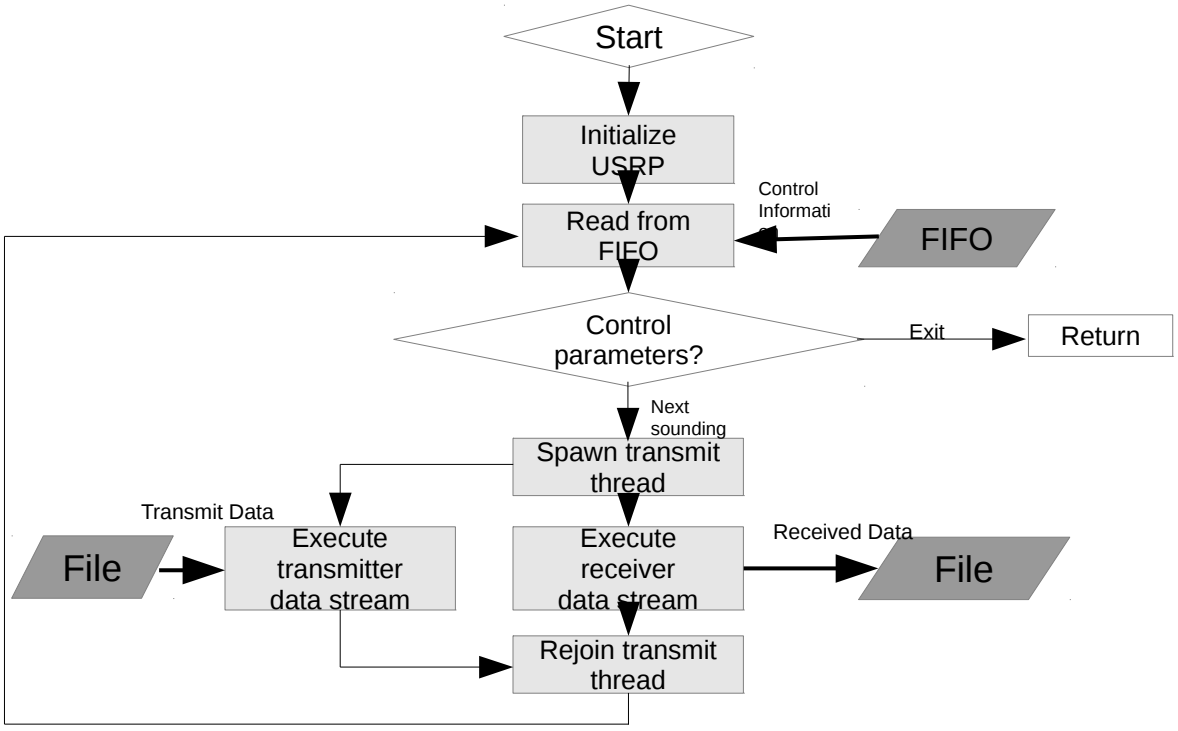

Figure 17: Flowchart for the sounder program

The first step in program execution is allocating memory for the program space and initializing the USRP device, which includes setting the clock reference for the transmit and receive streams. The UHD *multi_usrp* class has the methods *set_time_now()* and *set_time_next_pps()*, which are two different ways to set the clock reference. When the program executes, the *set_time_now()* method will set an arbitrary time for a particular USRP device. This means that the USRP reference clock will have a somewhat arbitrary time value assigned to it because the timing relation between the host computer and the USRP is indeterministic. This method would be invalid for setting the

time on multiple USRP devices since each one would have a slightly different reference time, but for a single USRP device the value assigned to a time event is arbitrary and the transmit and receive streams can be synchronized to the same arbitrary time event. The *set_time_next_pps()* would be a more appropriate method to synchronize multiple USRP's. Each device must have the same external reference clock and the same PPS input. The *set_time_next_pps()* method would tell each USRP device to set its internal clock reference to some value, all at the same PPS instant. In this way two or more USRP reference clocks can be synchronized with high precision.

The program *sounder* is a multi-threaded program, where the transmit and receive streams are executed in different threads. On a computer running the Linux operating system (OS), this means a separate process is spawned to handle the transmit stream and the operating system (OS) schedules access to the computer resources for the transmit and receive streams separately and the transmit and receive streams will not block each other's execution. Technically, this system is not perfect because the OS is not a real-time system and there is no guarantee that the transmit and receive processes will receive enough computing resources and be scheduled in a way to prevent transmit underflows and receive overflows. Other processes may compete for the computer's resources and prevent the critical processes from executing in time. Nevertheless, the speed and memory size of modern computers and the buffer size on the USRP are more than

adequate to essentially eliminate the possibility of overflow/underflow errors. Increasing the execution priority of the USRP functions, reducing the number of extraneous processes on the host computer, and increasing the TCP/IP buffer size on the host computer are additional precautions to reduce the possibility of error.

The program *sounder* is meant to be a process that runs fairly continuously, and so calling the program *sounder* for the first time involves no command line arguments. Instead, the program variables that define the RF center frequency, sample rate, transmit waveform file, and number of samples to collect are passed to the *sounder* process by means of a Unix named pipe (FIFO). This allows the time-consuming processes of memory allocation and USRP initialization to happen only once at program execution rather than at every sounding.

5.4 Waveform Generation

There are many viable software languages and tools for doing waveform generation. The Numerical Python (NumPy) package for Python 2.7 is chosen for this project for its popularity, ease-of-use, and free public license. Array manipulations, filtering, and type conversions are all simplified with NumPy.

The arguments passed to the waveform generation program *pulse_gen.py* are:

phase-code type, symbol rate, pulse repetition frequency, number of pulses, and sample rate. The program checks the arguments to ensure that there are an integer number of samples per symbol and symbols per pulse period.

The *pulse_gen* program begins with the array of phase code values defined by a lookup table in the program. For the 13-bit Barker code it is a 13-length array of values +1 or -1. Zeros are appended to the code to satisfy the pulse repetition period, the array is replicated to create the number of requested pulses, and the time series is over-sampled to achieve the desired sample rate.

Finally, the raw signal is low-pass filtered to remove out-of-band content. For this purpose the root-raised cosine (RRC) filter is used, which is popular in digital communications for both pre- and post-filtering of a signal because it satisfies the Nyquist criterion for eliminating inter-symbol interference. The combined response of the two filters is a raised-cosine filter, whose in-band and out-of-band spectral shapes are mirror images of each other. For the time-domain response, this translates to nulls at every integer number of symbol periods offset from zero and therefore minimizes possible inter-symbol interference caused by channel dispersion. The following figure illustrates the frequency response and impulse response of two cascaded RRC filters.

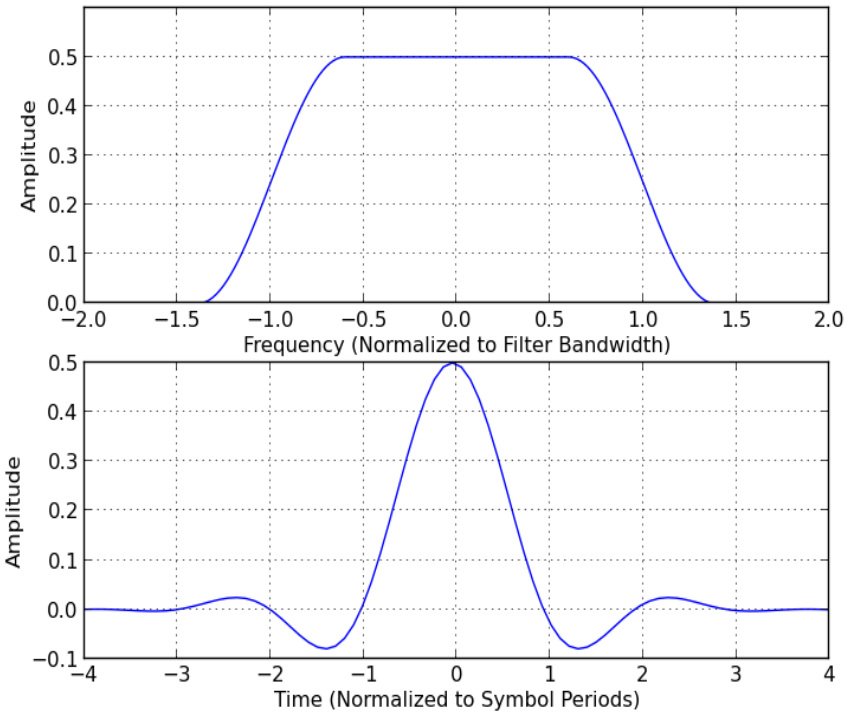

Figure 18: Frequency response and impulse response of raised-cosine filter

The transmission data sent over-the-wire to the USRP is an array of 16-bit signed integers, so the array floating-point values in the range -1.0 to + 1.0 is converted to integer format with range -32767 to +32767 and saved to a binary-type file, ready to be read and delivered to the USRP by the sounding program.

A transmit/receive (T/R) logic signal must be generated by the USRP and sent to the analog hardware for switching purposes. It must have a precise timing relationship with the transmit stream in order to maximize receive time, maintain transmit signal integrity, and prevent damage to sensitive receive components. The timing precision is

ensured by dedicating the least significant bit in the transmit stream as the T/R signal, and it is routed to General-Purpose Input/Output (GPIO) pin 15 on the USRP. The T/R timing is contained in the transmit stream, and so as long as the transmit data file is written correctly there is no possibility of mistiming the T/R signal. The figure below illustrates the filtered and unfiltered 13-bit Barker code, as well as the timing relationship between the transmitted pulse and the T/R signal.

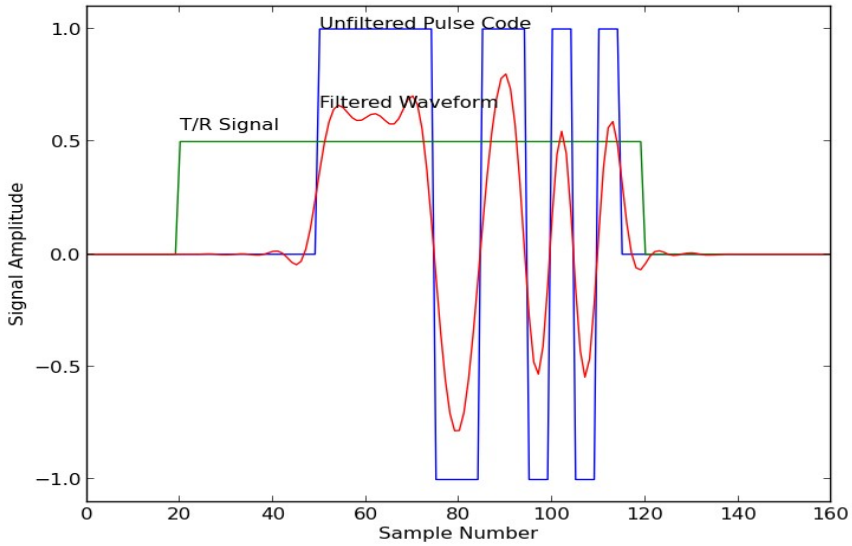

Figure 19: Transmit waveform with T/R signal overlayed

5.5 Processing the Received Signal

The complex received signal produced by the USRP is composed of 16-bit signed

integers, and it is converted to float type on the host computer to minimize numerical error. Then the received signal is correlated with the known transmitted pulse, a process known as matched filtering. The processing gain associated with this filtering is equal to the length of the pulse code, and it is illustrated below.

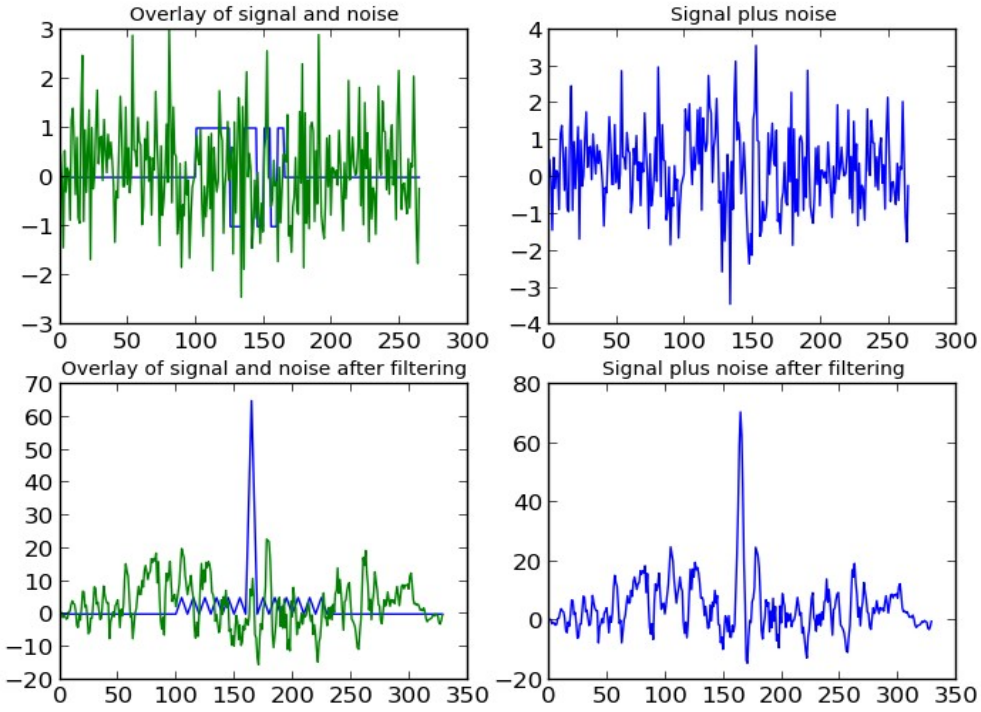

Figure 20: Demonstration of matched filtering

After matched filtering, the peak of the cross-correlation rises significantly from the surrounding noise.

Before and after the filtering, the signal is a one-dimensional vector of complex signal amplitudes spanning the entire time of the pulse sequence. In order to examine the

received signal time-series across multiple pulses, the array is reshaped into a range-Doppler grid.

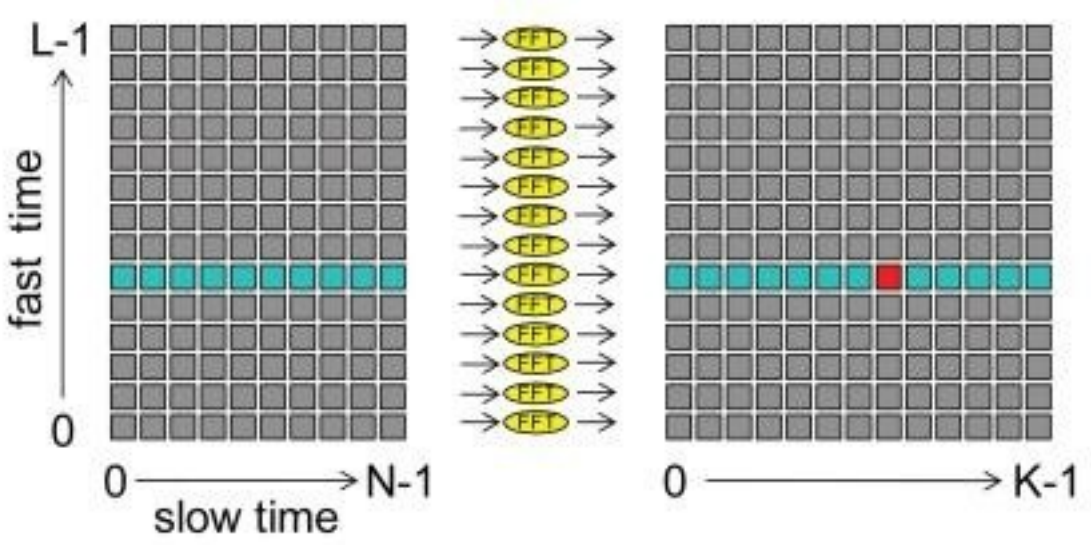

Figure 21: Range-Doppler binning

The one-dimensional vector is reshaped into a two-dimensional vector with *L* rows and *N* columns. The number of rows *L* is equal to the number of range bins in each pulse record, and the number of columns *N* is equal to the number of pulses in the sequence. The time-series of signal amplitudes in the slow time dimension (across pulses) is transformed to the frequency domain using the Fast-Fourier Transform (FFT) to find the peak signal amplitude which corresponds to the Doppler frequency of the returned signal in that range bin. The coding gain associated with Doppler processing is equal to the number of pulses in the sequence, and it is illustrated below.

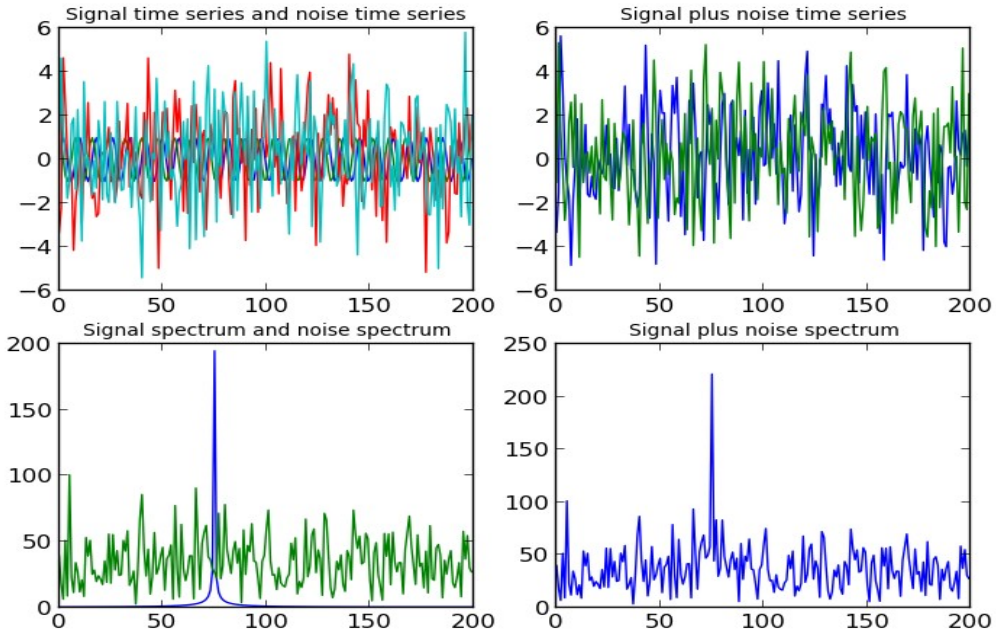

Figure 22: Demonstration of Doppler processing gain

Many advanced techniques exist for parameter-estimation of a signal time-series. For ionospheric sounding the demands are rather relaxed since it is assumed that only one spectral peak will exist in each range-bin, the number of time-series points is usually large, and none of the spectral moments need to be estimated with outstanding accuracy. The primary goal in analyzing the time-series is just testing for the presence of a significant signal, so processing of each range bin may be done simply by windowing, zero-padding, and applying the Fast-Fourier Transform (FFT). Computational efficiency of the Doppler-processing can be improved if some amount of unambiguous

Doppler velocity is sacrificed. Groups of consecutive slow-time columns are summed, resulting in a smaller number of slow-time columns and a smaller effective pulse-repetition frequency. The unambiguous Doppler velocity is reduced, but the number of points in the FFT is decreased, lessening the computational burden. For ionospheric sounding the pulse repetition frequencies are on the order of 200 Hz and the wavelengths are on the order of 100 meters, so the maximum unambiguous Doppler velocity is on the order of thousands of meters per second – much larger than the greatest expected velocity. Thus, it is reasonable to sacrifice the maximum unambiguous Doppler velocity to improve the computational efficiency.

To produce a sounding record at a single frequency, the power in each range bin is found by taking the maximum power level in each range bin's Doppler spectrum. Ranges that contain a coherent target will yield a significant spectral amplitude and a relatively large signal power for the corresponding range bin. An example of a single sounding record is illustrated below.

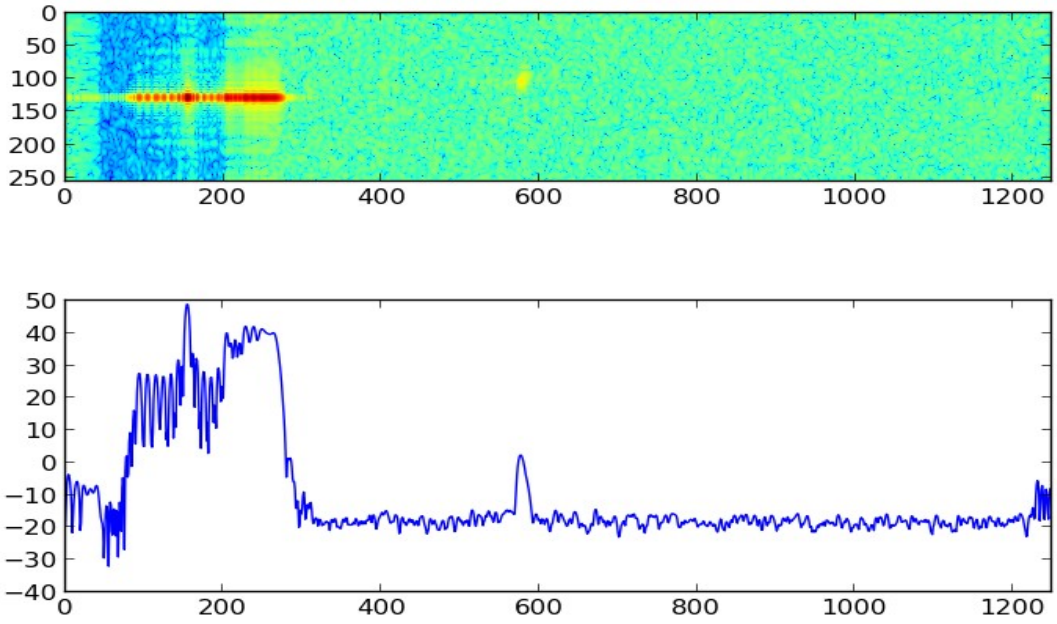

Figure 23: Example sounding record

The horizontal axis for the two figures is the fast-time dimension (range). The top figure shows power as a function of range and Doppler frequency, with the center row containing cells of zero Doppler velocity. The bottom figure is a plot of the maximum Doppler spectral power in each range bin. The left side of the figures show the transmitted pulse and switching transients, and near the center of the figures (range bin 590) there is evidence of a reflected target.

5.6 Scheduling

Scheduling refers to the sequence of discrete sounding frequencies that make up an ionogram image, and also when or how often each ionogram is created. This section will discuss primarily the choice of sounding frequencies that are used to produce an ionogram.

A single ionogram might involve almost 400 soundings, stepping through frequencies from approximately 0.5 MHz to 20 MHz in small 50 kHz increments. However, many of these soundings may yield redundant or obscured signal returns. For instance, many of the soundings may have a frequency greater than the plasma critical frequency and so no reflections are returned, or many of the soundings may fall on or near a competing broadcaster and the received data is noisy.

Rather than “blindly” stepping through a list of sounding frequencies, a more frugal method would utilize a-priori information to maximize performance and minimize expenses in time and energy. A-priori sources might include the noise spectrum in the HF band, recent sounding results, and temporal trends in ionospheric structure. Currently, the ionosonde described in this project uses only the received noise spectrum to guide the sounding schedule, but the software is organized in a way that anticipates the integration of more advanced algorithms to guide the sounding schedule. The flexibility

built-in to the system sets this ionosonde apart from most existing ionosondes by allowing for the development of novel and more efficient sounding schedules.

To implement the simple clear-frequency search, the program *choose_freq.py* starts by sampling the RF spectrum, performing the necessary averaging, and transforming it to the frequency domain. It finds the quietest frequency in each interval of a step size requested by the user, perhaps 100 kHz. For each 100 kHz interval, the 3 dB bandwidth of the "quiet" interval is determined, and the center of this quiet interval is the chosen sounding frequency. These sounding frequencies are written to a text file, along with their respective 3 dB bandwidths. This text file is read by the *sounder* process, and in this way the ionosonde chooses center frequencies and waveform bandwidths that are less likely to collide with existing noise sources and broadcasters.

The above procedure to locate quiet frequencies is simple but effective. More complicated algorithms to determine the best sounding frequencies may be integrated into the software stack in the future. The following figure illustrates the relationship between the various programs that make up the ionosonde software, where the program *choose_freq.py* may be altered or amended to incorporate more advanced scheduling algorithms.

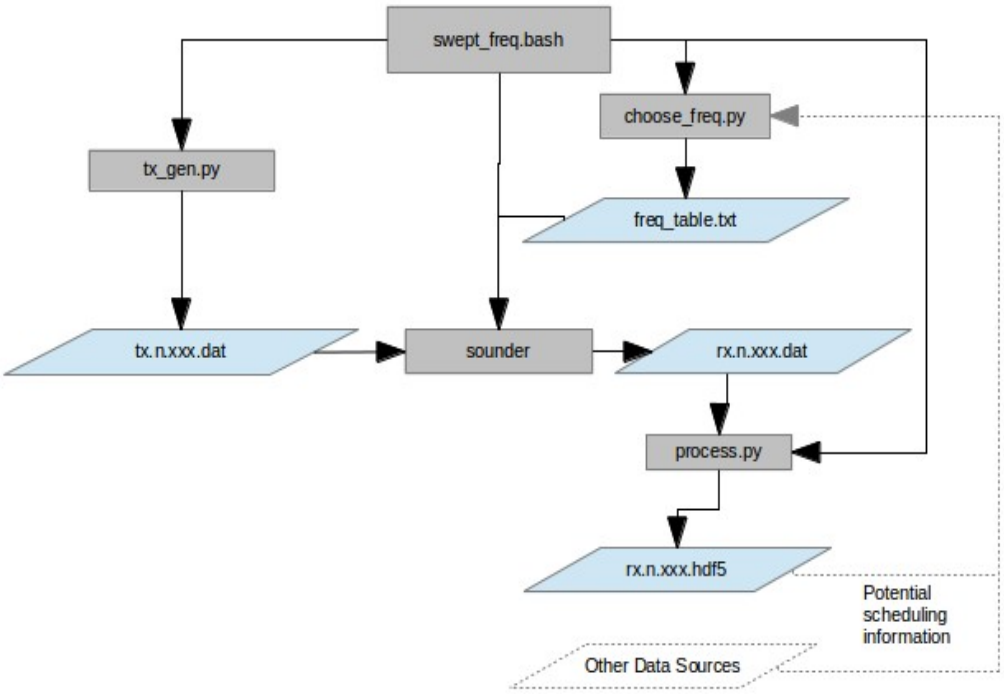

Figure 24: Sounding flow control

The script *swept_freq.bash* coordinates all other programs, first calling the *choose_freq.py* program to create a table of frequencies and bandwidths for sounding. The script then checks for the existence of the desired transmit waveform files, and if they do not exist it calls the *tx_gen.py* program to create them. The program *sounder* is called with the contents of the frequency table being piped to its standard input. When the entire received signal of a pulse sequence is written to a binary file, the program *process.py* is called, and *process.py* creates a single sounding record out of the raw file and adds it to an ionogram record in HDF5 file format. The program *sounder* continues to run until the frequency table is exhausted.

As the figure illustrates, there are other sources of information that could guide the sounding schedule besides the noise spectrum. A potential mode of operation might utilize information of the most recent sounding records to determine the next frequencies to use for sounding, dynamically constructing the frequency table during the course of a single ionogram. For instance, the logic in *choose_freq.py* could estimate a value of the plasma critical frequency based on relatively few soundings. After the critical frequency is known to within some certainty, the ionosonde may save time and energy by not operating on frequencies above the critical frequency where no reflected signal is expected. Similarly, a large number of soundings may be done at or near the predicted frequency cusps corresponding to the D, E, and F layers of the ionosphere, maximizing the resolution of the ionogram for these important zones. The architecture of the sounder's software is flexible and allows these and many other unexplored scheduling algorithms to be easily integrated with the system.

The following table describes the basic functions of each program. The description of each program's function is non-specific, again emphasizing the fact that future pieces of software can be easily integrated with the existing structure.

Table 1: Summary of software programs

Program	Description	Inputs	Outputs
swept_freq.bash	This script is called to initiate the sounding process. It calls other programs and handles inter-process communication.		
sounder(.cpp)	Compiled c++ program using the UHD API that handles all interaction with the USRP device. Initializes the USRP and handles transmit and receive streams.	Center frequency, sample rate, number of samples, transmit waveform file	Received data (raw) file
tx_gen.py	Python script using the NumPy package that creates the entire transmit waveform and writes	Number of pulses, symbol rate, sample rate, pulse repetition frequency	Transmit waveform file
process.py	Python/NumPy script that processes the raw received data. Filters the signal, correlates the signal with the transmit pulse sequence, and calculates the Doppler spectrum for each range bin.	Number of pulses, symbol rate, sample rate, pulse repetition frequency	Received data (processed), ionogram in HDF5 format
choose_freq.py	Python/NumPy script that selects the best sounding frequencies based on noise power in the HF spectrum.	Minimum and maximum frequencies, step size	Sounding channel(s)

5.7 Completing the System

The focus of the design is on properly transmitting and receiving signals in the digital domain using computer software. To actually perform soundings and evaluate the quality of the system, an analog frontend and HF antenna are required. Since the signal processing chain of the ionosonde allows for large gains in SNR, results are achievable with small output power and only modest gain from the antenna. A very basic transceiver was constructed for early tests, and later a Super Dual Auroral Radar Network (SuperDARN) transmitter was borrowed and incorporated for testing.

A system block diagram below illustrates the basic components of the analog

front-end: switch, transmit and receive amplifiers, and bandpass filter.

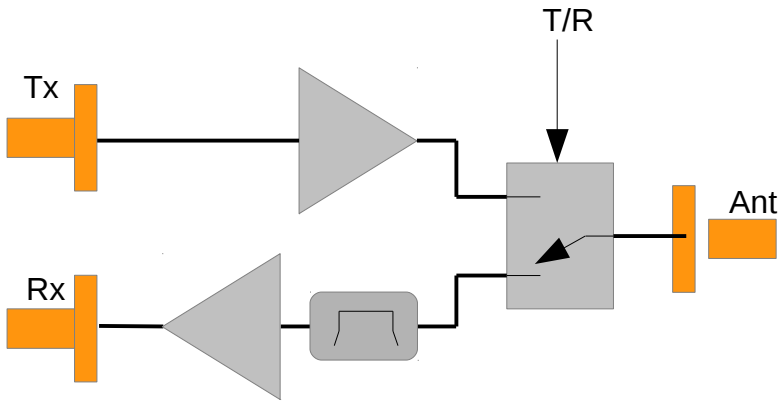

Figure 25: Analog front-end

In monostatic radars both the transmitter and receiver are connected to a single antenna using some kind of duplexer. In the microwave bands a ferrite circulator is often used, but for the long wavelengths in the HF band only a switch is practical. Switch times of only a few microseconds are desirable, so a solid state switch made using PIN diodes or MOSFET transistors is necessary. In this project a MiniCircuits ZX80-DR230-S switch was used, driven by a T/R signal from one of the USRP's GPIO pins.

A transmit amplifier drives the antenna, increasing the amount of power that will be reflected off the ionosphere and received by the ionosonde. There are numerous choices of amplifiers that are suitable for this frequency range. The MiniCircuits ZFL-1000-VH general purpose amplifier was used for this project, which has 20 dB of gain and a 1 dB compression point of 25 dBm.

An amplifier chain in the receive path is necessary to use the full range of the ADC. High linearity of the receive chain is important, while low noise figure is not critical. Two MiniCircuits ZFL-1000-VH amplifiers were cascaded for a net gain of 40 dB. The linearity is specified as +38 dBm IP3. To limit possible non-linear effects in the receive chain and ADC, a band-pass filter was used to reduce the total power at the receiver front-end. A custom 5th-order high-pass filter at 3 MHz was cascaded with a Mini-Circuits SLP-30+ 30 MHz low-pass filter. The 3 MHz to 30 MHz passband had significantly lower power than that of an unfiltered spectrum, particularly because the AM broadcast bands below 2 MHz and FM broadcast bands around 100 MHz were strongly attenuated.

For the second round of testing the above basic transceiver was supplemented with a SuperDARN transmitter, made by Scientific Instruments Limited (SIL) for the Mid-Scale Infrastructure (MSI) series of radars. Only the first amplifier stage of the transmitter was used in testing, but the 10 W of peak output power was a 13 dB improvement over the 500 mW used in the first round of testing.

One of the largest obstacles in implementing an ionosonde is constructing a vertical-gain antenna suitable for all wavelengths roughly from 300 meters to 10 meters (1 MHz to 30 MHz). A large wire antenna mounted on one or more towers with size the same order of magnitude as the longest radio wavelength would be the ideal solution, but

smaller antennas are possible where certain performance parameters can be sacrificed for smaller size and expense. There is a fundamental trade-off in antenna design between physical size, quality factor (bandwidth), and efficiency [Chu, 1948].

For the first round of testing, a small tuned magnetic loop antenna was used, which greatly limits bandwidth but imposes no theoretical limits on efficiency. The magnetic loop antenna is an electrically small antenna with small radiating resistance. It is best implemented as the inductive component in a series LC resonant circuit, which maximizes the current per cycle and thus the radiated power. The losses in the resonant circuit come from radiating resistance of the antenna and ohmic losses in the capacitor and the loop itself. The radiation resistance of most loops is on the order of mΩ, so any small resistive losses in the circuit are significant compared to the radiation resistance and greatly affect the efficiency of the antenna. The small resistance (both radiating and ohmic) makes for a large quality factor. Q values are on the order of 100, so bandwidth values for 3 MHz are on the order of 30 kHz. This is suitable for sounding on only a single frequency, and changing the sounding frequency requires changing the resonant frequency F by varying the capacitance C that resonates with the loop inductance L according to:

$$F = \frac{1}{2\pi\sqrt{LC}} \quad . \tag{5.1}$$

The antenna is series-resonant at the operating frequency, and the input impedance is small and purely resistive. Coupling to the antenna is done via a 7:1 voltage transformer that increases the secondary impedance by a factor of 49 and acts as a balun between the balanced antenna and the unbalanced coaxial feedline. The magnetic loop antenna is illustrated in the figure below.

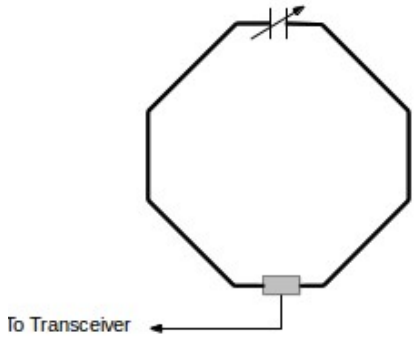

Figure 26: Magnetic loop antenna

The measured standing wave ratio (SWR) is plotted in the following figure.

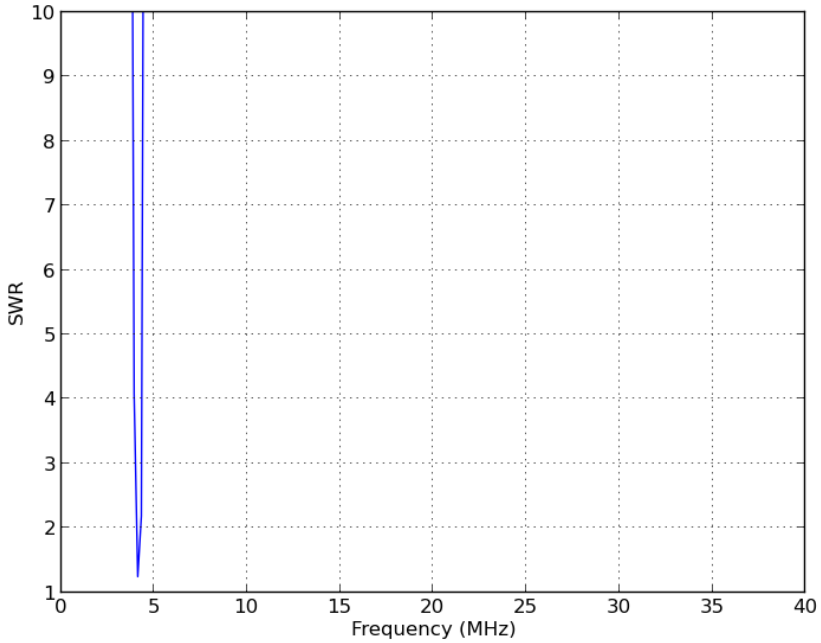

Figure 27: Measured standing wave ratio of magnetic loop antenna

The SWR shows good matching at a single frequency and narrow bandwidth. The antenna is essentially useless at all other frequencies.

For the second round of testing, an offset-fed dipole was used. The regular dipole antenna is highly efficient, but it is resonant only at its fundamental frequency and at odd harmonics of the fundamental. The offset-fed dipole is a variation to the regular dipole antenna, where instead of each leg of the dipole being equal in length, the legs have a length ratio of 2:1. This allows the offset-fed dipole to have a matchable feed-point impedance at the fundamental and at both the even and odd harmonics of the fundamental. It is a somewhat broadband antenna suitable for modest attempts at

ionospheric sounding.

The feed-point impedance is roughly 200 Ω, so matching to the 50 Ω feedline was done using a 2:1 voltage transformer/balun. The antenna was constructed of 14-AWG copper wire and mounted approximately 3 meters off the ground. Its total length was 75 meters. The offset-fed dipole is illustrated in the figure below.

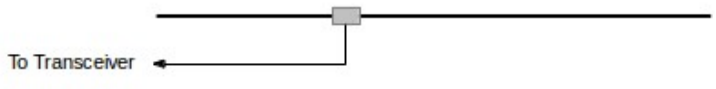

Figure 28: Offset-fed dipole antenna

The following figure is the measured SWR of the antenna from 1 MHz to 40 MHz.

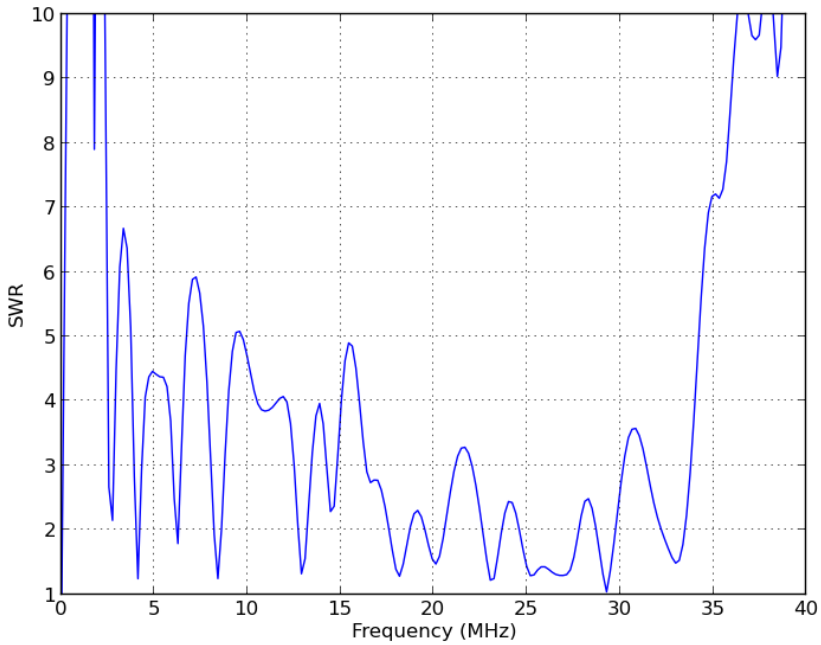

Figure 29: Measured standing wave ratio of offset fed dipole antenna

The figure shows excellent matching at several frequencies in the band of interest from 1 MHz to 20 MHz. The fundamental resonance is evident at approximately 2.0 MHz and the harmonic resonances are evident at 4.0 MHz, 6.0 MHz, and 8.0 MHz. Between these frequencies the matching is poor, but the antenna is still usable.

Chapter 6 Sounding Results

6.1 Single Frequency Soundings

The first trials of the ionospheric sounder prototype involved a fixed frequency of 4.15 MHz and integration times of almost 20 seconds for each sounding. The basic 500 mW transceiver was used with the simple magnetic loop antenna. Soundings were performed every minute on the days of November 25 and 26, 2013. The following figure displays the results.

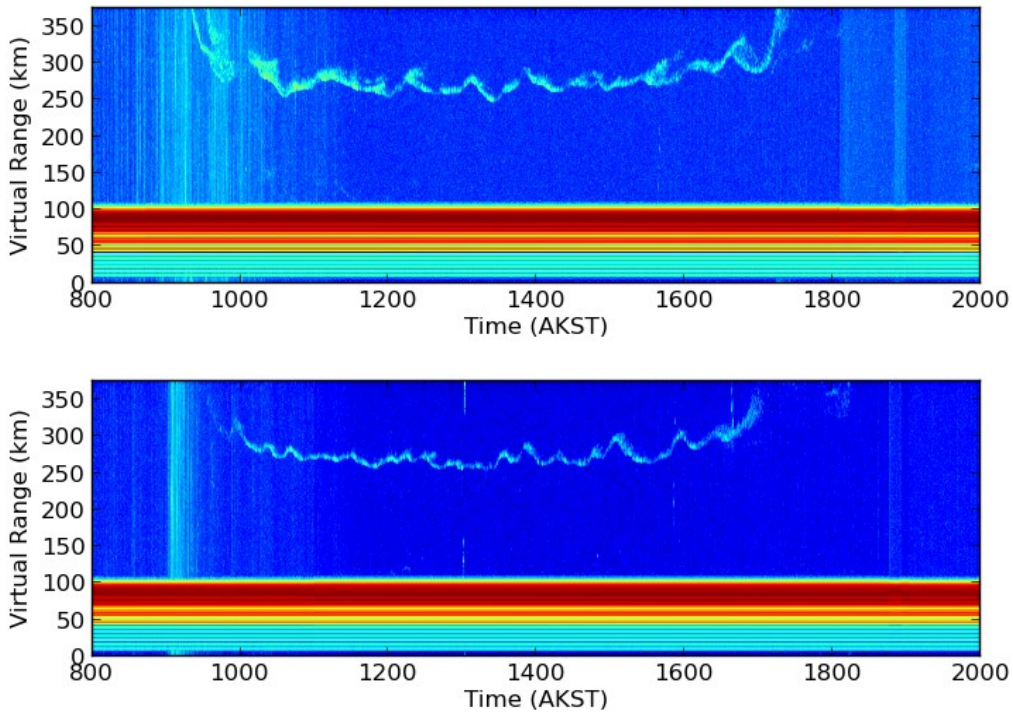

Figure 30: Example fixed-frequency sounding results

The results show signal-to-noise ratios of over 12 dB for many soundings with some soundings exceeding 20 dB. The expected behavior of the ionosphere is recorded with small, roughly 30 km variations in the height of the ionosphere during daytime and then a rapid recombination of the atmospheric ions in the several hours after sunset. The first results are encouraging given that the peak transmit power is less than 500 mW and the antenna gain in the vertical direction is estimated less than -10 dB.

The noise power at the receiver was compared to the expected noise power for a semi-urban environment, and the received noise spectrum was compared to the previous

assumptions that the noise is white. The following figure is the measured noise spectrum at the input to the receiver.

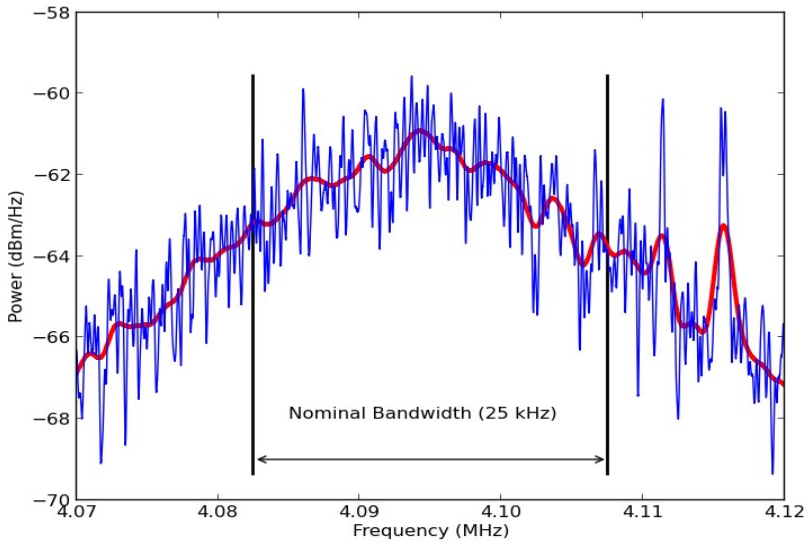

Figure 31: Color of actual received noise spectrum

The signal spectrum at the front end of the USRP receiver is a product of atmospheric noise plus interference, the antenna transfer function, analog filtering, and the receiver amplifiers. The peak spectral power at the receiver is roughly -60 dBm/Hz for 60 dB of gain in the receive amplifiers. This corresponds to -120 dBm of noise power, which is 54 dB greater than the thermal floor. A more efficient antenna would pull in perhaps another 10 dB of power, giving a rough estimate of the atmospheric noise power to be 64 dB above thermal noise. This value is consistent with what the noise tables predict for a semi-urban environment at this frequency.

The received spectrum contains many narrow-band power sources, which defies the assumption of white spectral power. However, taking a closer look at the spectrum within the bandwidth of a single sounding, there is evidence of only weak interferers compared to the broadband noise level, and the noise spectrum is approximately flat when considering the non-flat transfer function of the antenna.

6.2 Swept Frequency Soundings

The antenna was changed from a magnetic-loop type antenna to an offset-fed dipole antenna, and soundings were performed on frequencies between 3 MHz to 10 MHz on February 5, 2015 at 19:15 AKST. Seventy discrete frequencies were used, each having an integration time of 5.5 seconds.

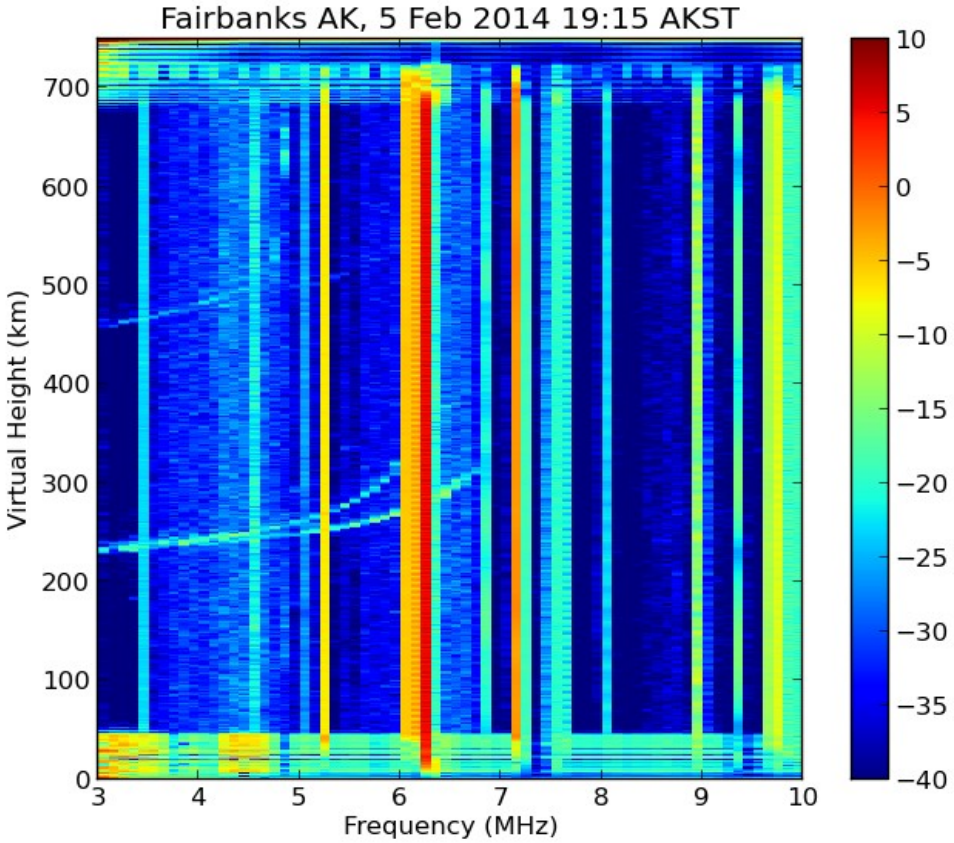

Figure 32: Swept-frequency ionogram image

The result is typical of ionograms for the frequencies above 3 MHz. The o-mode and x-mode traces are both evident, and the second hop can be seen at approximately twice the virtual height of the first hop. The foF2 frequency can be estimated to be between 6.5 MHz and 7.0 MHz.

In the final round of testing, the basic clear-frequency search algorithm was

implemented and the basic transceiver was supplemented with a SuperDARN transmitter. The offset-fed dipole antenna was lifted as high off the ground as possible and fine-tuned for minimum reflection coefficient to the transmitter.

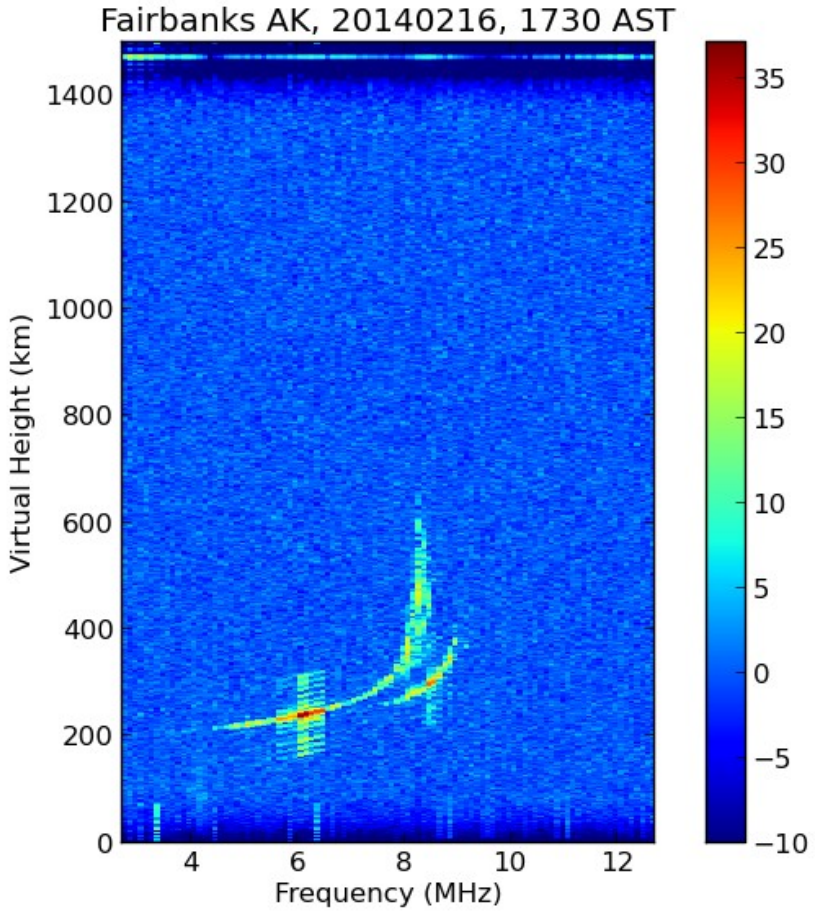

Figure 33: Swept-frequency ionogram produced with improved analog components

The signal to noise ratio is improved, and the ionogram traces are more clearly defined

than in the previous test. An unfortunate consequence of good signal-to-noise ratio is that the side-lobes of the matched-filter function are clearly visible above and below the true ionogram trace. Nevertheless, in this case the imperfect autocorrelation function does not obscure a neighboring trace, and locations of the o- and x-mode traces are both easily recognizable.

Chapter 7 Conclusion

7.1 Evaluation of Performance

Overall, the performance of the prototype ionospheric sounder exceeded expectations. For most days the SNR at the receiver exceeded 12 dB and sometimes was in excess of 20 dB, even for an inefficient antenna and a peak transmit power of only 500 mW. With a higher-power transmitter and an antenna that meets the demands for bandwidth and gain, there is little doubt that this design for an ionospheric sounder would reliably produce good data.

The following table lists the important performance specifications of the prototype ionospheric sounder.

Table 2: Summary of performance parameters

	Minimum	Maximum	Typical
Receiver Dynamic Range (dB)	N/A	N/A	72
Peak Transmit Power (W)	0.5	10	10
Average Transmit Power (W)	0.05	1	1
Matched Filter Gain (dB)	0	14	12
Doppler Processing Gain (dB)	0	33	21
Range Resolution (km)	1*	6	6
Maximum Range (km)	250	2000	1000
First Range (km)	20	6	78
Maximum Doppler (Hz)	25	100	50
Doppler Resolution (Hz)	0.1	10	1

7.2 Costs

This project describes an inexpensive setup for performing ionospheric sounding. Baseband signal processing is performed digitally using a general-purpose computer, the digital front-end is the Ettus Research USRP, and the low power requirements allow for inexpensive components in the analog front-end. The following table lists expenses, minus an HF antenna. Since the exact cost of equipment depends somewhat on choice of components, a minimum and maximum estimated cost are listed.

Table 3: Summary of expenses

Item	Estimated Cost (USD)	
	Minimum	Maximum
Computer	400	1500
USRP	800	2200
Analog Frontend	500	3000
Totals	**1700**	**6700**

The choice of computer and USRP affect sounder performance only minimally, and rather depend on the user's network configuration, desired peripherals and features for the computer, and the need for extraneous space on the USRP's FPGA. The cost of the computer and USRP in the prototype ionosonde were $400 and $1850, respectively. The exact cost of the analog frontend depends on desired performance level. Greater transmitter power allows ionograms to be completed in less time, but comes with greater

cost of the transmitter itself and greater costs in increasing the power-handling ability of the other circuit components in the transceiver. Better front-end filtering improves the dynamic range of the receiver and thus its sensitivity, but requires more circuit components, and perhaps a complex tunable filter. Other features of a transceiver, such as adjustable receiver gain, SWR metering, remote turn on/off, fault detection, etc. increase the complexity and cost of the transceiver. The analog front-end of the prototype ionosonde was composed entirely of connectorized MiniCircuits components, with a total retail price of $940. The performance of the prototype transceiver could be replicated with surface mount devices and custom PCB fabrication at a fraction of the cost in parts.

7.3 Future Improvements

There are two features common in modern ionosondes that are not implemented in this design, particularly o-mode/x-mode discrimination and angle-of-arrival determination. The first feature is a valuable, perhaps necessary, data product for inverting ionograms into electron content information and can be implemented with the addition of a second receive channel and some additional signal processing. Angle-of-arrival information provides supplemental data to the basic ionogram that is less valuable,

less reliable, and requires a relatively large investment in additional electronics and additional antennas.

The o-mode and x-mode can be discriminated by observing the polarization of the reflected wave, which will be left-hand circularly polarized for the o-mode and right-hand circularly polarized for the x-mode in the Northern Hemisphere. A system with two antennas that are offset 90 degrees in space and 90 degrees in phase whose signals combine will select for either the o-mode or x-mode, depending on whether the 90 degrees phase offset is leading or lagging. Each of the two received signals are mixed to baseband, filtered, and down-sampled in the USRP. The host computer may then apply the 90 degree phase shift to one of the two signals and calculate the sum of the two to determine one polarization and calculate the difference between the two to determine the other polarization.

Angle-of-arrival estimation is performed by summing the coherent signals from spatially-distributed receive antennas. By manipulating the phase delay on each of the received signals it is possible to increase the net antenna gain in a certain azimuth/elevation angle direction and thus determine from which part of the sky signal reflections originate. The number of discrete beam directions depends on the desired angular resolution and the amount of signal processing burden willing to be incurred. Since the number of receive antennas would be small, perhaps only four, the beam width

of the antenna array would be relatively large. Angle-of-arrival estimation is fundamentally limited by the receive antenna beam width, so would make sense to only calculate the received power at a small number of beam angles, greatly reducing the signal processing burden without any loss of information.

7.4 Deploying a Terrestrial Ionosonde

It has been demonstrated that even the simplest hardware and a crude magnetic loop antenna are able to produce ionospheric sounding results. A system that would be put into service for scientific purposes warrants a greater investment in equipment, especially for the antenna. High vertical gain HF antennas include the inverted delta, rhombic, and log-periodic antennas. The following figure illustrates a possible low-cost antenna tower that could be placed on University of Alaska property.

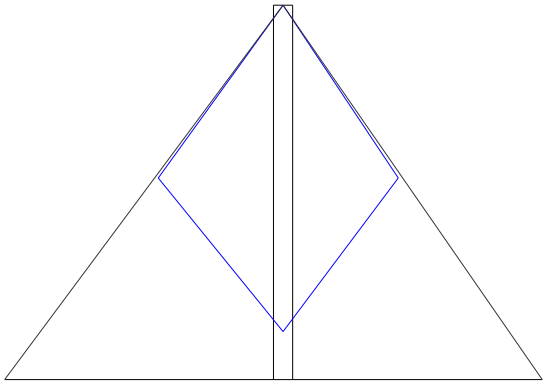

Figure 34: Proposed ionosonde antenna for local installation

The antenna is a 100 foot tall rhombic-type antenna, composed of small gauge stranded copper wire, driven at the bottom with a 2.5:1 balun and terminated at the top with a 600 Ω resistor. The antenna wires are secured at the top and bottom to a steel lattice tower, and secured on the left and right ends to Kevlar guy wires using crimped hog rings. The antenna tower requires only a small, shallow concrete foundation, four ground anchors for guy wires, and can be assembled manually without a crane or man-lift. By NEC computer simulation, the antenna provides between 5 dB and 8 dB of vertical gain and the SWR is less than 3 for all frequencies between 1 MHz and 20 MHz.

Radar electronics may be housed in a small weather-proof cabinet, or in a nearby building with 120 VAC power and an Internet connection. Ideally, the antenna should be located within several hundred feet of a pre-existing building.

The cost of a 100 ft. Rohn tower is $7,000 and the cost of Kevlar guys, turn-buckles, and copper wire is approximately $1,000. The cost of installation by a local

contractor, including preparing a concrete pad and driving ground anchors, is estimated to be $2,500, making the total cost of the antenna approximately $10,500. It is assumed that the antenna can be installed near a pre-existing building that can house the electronics and provide power and networking.

7.5 Deploying a Space-Borne Ionosonde

In recent years the costs of developing spacecrafts and launching them into Earth orbit has fallen, particularly with the advent of “cube-sat” satellites, a standard which describes satellites whose physical dimensions are units of 10 cm cubic sections and which is promoted by many international organizations and universities and supported by NASA. To deploy a space-borne sounder as described in this thesis would require aggressive improvements in electronic circuitry and antenna design for reliability in space and compact size. Nevertheless, this project describes a design for an ionospheric sounder whose simplicity, low power consumption, flexibility, and high efficiency would be advantageous in a space environment.

To port the ground-based sounder design to a space-borne sounder would require custom board fabrication that would integrate the functionality of the host computer, the USRP digital front end, and the analog frontend transceiver into a single compact

package. Critical digital components include a general-purpose microcomputer, communications bus, RF oscillator, DAC and ADC, and FPGA. Critical analog components include forward transmit amplifier, receive amplifier, bandpass filter, and duplexer. The functionality of the entire system can be described simply when broken into discrete components, but the time and effort involved in assembling a single-board solution should not be underestimated. Wherever possible, the design solution should draw on existing circuit schematics and layouts, such as those of the open-source hardware Ettus Research USRP, manufacturer-suggested layouts for single-board computers, etc. A space-capable design may require more strict specifications for components in regards to minimum and maximum temperature ratings and idle power consumption.

The scheduling algorithm for space-borne sounding may also be different than ground-based sounding. There are several key differences. The first difference is that the space environment is generally much less noisy than the ground environment at HF. This would allow for more modest integration times for each sounding. The second difference is that the distribution of plasma density in the topside of the ionosphere is simpler and more gradual. Rather than several distinct layers that exist on the bottomside of the ionosphere, the density of the topside ionosphere simply decays exponentially with altitude, and the spacecraft itself actually orbits in a low-density plasma. Therefore, the

frequencies of interest are different for space-borne sounding, and different operating modes are possible since the sounder can resonate with the immediately-surrounding plasma. Finally, the greatest difference is the velocity of a space-craft compared to a stationary ground-based sounder. A satellite in low-earth orbit travels at approximately 7 km/s, so it is desirable to complete an entire ionogram as quickly as possible. A spatial resolution of 100 km is desirable, so a complete ionogram should be completed in approximately 15 seconds. To achieve this might mean a coarser ionogram that sacrifices some frequency resolution to achieve better time (spatial) resolution.

References

Bullet, T. 2008. *Station Report: A New Ionosonde at Boulder.* Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, viewed 24 January 2014. <http://www.ips.gov.au/IPSHosted/INAG/web69/2008/boulder_vipir.pdf>

Chen, F. 2006. *Introduction to Plasma Physics and Controlled Fusion.* Springer, New York, NY.

Chu, L.J. 1948. *Technical Report No. 64: Physical Limitations of Omnidirectional Antennas.* Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA.

Ettus Research. (2013, September 10). "UHD Documentation." Retrieved February 12, 2014, from http://files.ettus.com/uhd_docs/doxygen/html/index.html

Ettus Research. (2014, February 07). "UHD-USRP2 and N2x0 Series Device Manual." Retrieved February 12, 2014, from http://files.ettus.com/uhd_docs/manual/html/usrp2.html

Levanon, N. 1988. *Radar Principles.* Wiley, Hoboken, NJ.

Piggot, W.R., & Rawer, K. 1972. *U.R.S.I Handbook of Ionogram Interpretation and Reduction.* World Data Center A for Solar-Terrestrial Physics, NOAA, Boulder, CO.

Rawer, K. 1993. *Wave Propagation in the Ionosphere.* Kluwer Academic Publishers, Norwell, MA.

Reinisch, B. 2007. *The Digisonde Portable Sounder – DPS Technical Manual.* University of Massachusetts Lowell Center for Atmospheric Research, Lowell, MA.

Richards, M., Scheer, J., & Holm, W. 2010. *Principles of Modern Radar: Basic Principles.* Scitech Publishing, Inc., Raleigh, NC.